University of Adelaide
Teletraffic Research Centre
Department of Applied Mathematics
PhD Thesis

Signalling in Product Form Queueing Networks

Bruce S. Northcote B.Sc.(Ma) Hons.
December 1993

Contents

List of Figures **iv**

List of Tables **v**

Abstract **vi**

Declaration **vii**

Acknowledgement **viii**

1 Introduction **1**

1.1 Motivation 1

1.2 Literature Overview 6

1.2.1 Characterisation of Product Form Queueing Networks 6

1.2.2 Signals: Negative Customers, Triggers and Interruptions 13

1.3 Outline 17

2 Batch Destruction of Customers **19**

2.1 System Description 21

2.2 The Equilibrium Distribution 22

2.3 Partial Balance 34

2.4 Gelenbe's Results 36

3 State Dependent Signalling 39

3.1 System Description 42

3.1.1 Customer and Signal Emission 43

3.1.2 Signal Conversion at the Destination Node 44

3.2 The Equilibrium Distribution 46

3.3 Special Cases 55

3.3.1 State Space Truncation 55

3.3.2 Network Partitions and Multiple Customer Classes 56

3.4 Manufacturing, Supply and Demand Models 59

3.5 Predator-Prey Models 62

3.6 Networks of Customer and Resource Queues 65

3.7 Communications Protocols 67

3.7.1 Once Only Verification Protocol 67

3.7.2 Repeated Verification Protocols 69

4 Quasi-reversibility 71

4.1 Quasi-reversible Processes 72

4.2 Networks of Quasi-reversible Nodes 74

4.3 Queueing Networks with Batches of Negative Customers 77

4.4 Triggered Linking of Quasi-reversible Nodes 81

5 Forced Batch Movement **90**

5.1 System Description 92

5.2 The Equilibrium Distribution 94

5.3 Applications 102

5.3.1 Stochastic Petri Nets with Constant Firing Rates 102

5.3.2 Resource Allocation Problems 109

5.3.3 Loss Networks 112

Concluding Remarks **117**

Bibliography **119**

List of Figures

3.1 Once only verification protocol 68

3.2 Repeated verification protocol 69

5.1 An example of a transient routeing chain 100

5.2 An arc that represents forced batch movement 101

5.3 Delayed forced transition firing 104

5.4 Forced transition firing with conflict resolution 105

5.5 Simulated net using conflict resolution 107

5.6 A typical loss network 112

5.7 Resource allocation model of a loss network 114

List of Tables

4.1 Parameter equivalence between coupled and Gelenbe networks 80

4.2 Parameter equivalence between coupled and batch negative networks . 80

5.1 Simulated and Exact Transition Enabling Probabilities 109

Abstract

Product form queueing networks are often used to model systems as diverse as communications protocols, computer systems, ecological systems, neural networks and manufacturing processes. For this reason a lot of effort has been invested by researchers in extending the set of queueing networks which are known to have a product form queue length distribution.

Recently Gelenbe demonstrated that queueing networks can include entities which he described as negative customers, triggers and signals and retain product form. This thesis describes and solves new networks (which include these entities) incorporating state dependent firing rates, multiple customer classes, batch movement and batch destruction. A number of applications for these networks are presented, including an application in which the queueing network results are applied to stochastic Petri nets to obtain a new class of nets which have product form.

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

SIGNED: ....

DATE: ....

Acknowledgement

I would like to thank Nigel Bean for the hours that he spent proof reading this thesis, and the many helpful suggestions he gave.

Special thanks must go to my supervisors, Bill Henderson and Peter Taylor, who, although being constantly overworked, were always ready to provide assistance, even late on a Friday afternoon.

This brings me to my last note of gratitude: To the staff and students of SPAM, and (especially) the staff of the University of Adelaide Club, many thanks for making postgraduate life so enjoyable.

Chapter 1

Introduction

1.1 Motivation

In real life we are constantly queueing, whether it be to buy tickets for a football game, or while waiting for our food to be served in a restaurant. Naturally, queueing is not restricted to people. Resources, for example, may queue at several stations as they pass through a manufacturing process. In abstract terms, a queue consists of an ordered set of customers (people, resources, etc.) waiting to be served at a station or node. The exact manner in which the customers are served depends on the service discipline of the queue. In a queue operating under the first come first served (FCFS) service discipline, customers are served one at a time in the order in which they arrive. Other service disciplines may serve the customer which is the most recent arrival first, or may serve many customers simultaneously. To describe fully the structural properties or dynamics of a queue we need to define the arrival stream of customers to the queue, the service discipline of the queue, and the amount of service that each customer requires. Because most real life arrival and service processes are stochastic, a good queueing model needs to include stochastic elements to take account of this randomness.

The mathematical model used to describe a queue consists of a node or station and stochastic processes that describe the arrival and subsequent service of customers.

Associated with a node is a queue length and a service discipline. Sometimes the ordering and types of customers within the queue also needs to be considered. With knowledge of the underlying stochastic processes it is possible to derive performance measures of interest for the model. These include, for example, the average queue length, mean waiting times of customers, and throughput of the queue.

A queueing network may be formed from a number of nodes by coupling them together to allow customers to move from one node to another (or leave the network entirely). This requires a routeing rule to be defined that specifies the way in which customers may move between queues. Customers in a network may arrive to a specific queue from outside the network, as defined by the arrival stream of the queue, or from another queue in the network, so that the input of a particular queue consists of some portion of the total input to the network and a portion (defined by the routeing rule) of the outputs of the other queues in the network. The state of a network of N queues at time t is described by a vector $\mathbf{n}(t) = \big(\mathbf{c}_1(t), \mathbf{c}_2(t), \ldots, \mathbf{c}_N(t)\big)$ where $\mathbf{c}_i(t)$ (possibly a vector) describes the configuration of customers in the queue at node i at time t. Note that if there is no need to distinguish between the customers in a queue then $\mathbf{c}_i(t) = n_i(t)$, where $n_i(t)$ is the number of customers in the queue at node i at time t. The state space of a network is defined to be the set S of all possible states $\mathbf{n}$ in which the network may find itself. The transitions of a network of queues are usually described by a set of transition probabilities (in discrete time) or intensities (in continuous time) $q(\mathbf{n}, \mathbf{n}')$. In discrete time these probabilities have the form

$$q(\mathbf{n}, \mathbf{n}') = Pr\Big(\mathbf{n}(t+1) = \mathbf{n}' \Big| \mathbf{n}(t) = \mathbf{n}\Big), \quad \forall \mathbf{n}, \mathbf{n}' \in S, \ t \geq 0.$$

$q(\mathbf{n}, \mathbf{n}')$ is the probability that the state of the network at the next discrete time point will be $\mathbf{n}'$ given that its current state is $\mathbf{n}$. In continuous time the transition intensities are

$$q(\mathbf{n}, \mathbf{n}') = \lim_{s \to 0^+} \frac{Pr\Big(\mathbf{n}(t+s) = \mathbf{n}' \Big| \mathbf{n}(t) = \mathbf{n}\Big)}{s}, \quad \forall \mathbf{n}, \mathbf{n}' \in S, \ t \geq 0.$$

In the rest of this thesis, unless otherwise stated, we will only consider the continuous time case.

The stochasticity of the arrival and service processes implies that the queue lengths in a network will fluctuate over time with a degree of randomness. The transient (time dependent) queue length distribution is

$$\pi(\mathbf{n}, t) = Pr\Big(\mathbf{n}(t) = \mathbf{n}\Big).$$

This distribution may be found, in discrete time, by solving the Kolmogorov equations

$$\pi(\mathbf{n}, t) = \sum_{\mathbf{n}' \in S} \pi(\mathbf{n}', t-1) q(\mathbf{n}', \mathbf{n}), \ \ t \geq 0.$$

Due to the complexity of queueing networks it is usually extremely difficult to solve these equations and obtain an expression for $\pi(\mathbf{n}, t)$.

A stable network may be loosely described as one in which the servers operating at each node (in a manner described by the service discipline) are able to cope with the service demands placed upon them by arriving customers. Statistical equilibrium is achieved in a stable network when the network no longer changes its characteristics under the stochastic influences of the arrival and service processes. It follows that the equilibrium (time independent) queue length distribution will be simpler to obtain than the transient queue length distribution. Fortunately, many queueing networks approach statistical equilibrium quickly, ensuring that a study of the equilibrium distribution will provide sufficiently accurate information about the behaviour of the network. There are of course many classes of queueing networks which do not reach equilibrium quickly, but even for these networks a study of equilibrium behaviour is beneficial as it will often reveal the intrinsic character of the network. For a stable queueing network the equilibrium distribution is

$$\pi(\mathbf{n}) = \lim_{t \to \infty} \pi(\mathbf{n}, t),$$

and it satisfies the global balance equations (GBEs)

$$\pi(\mathbf{n}) = \sum_{\mathbf{n}' \in S} \pi(\mathbf{n}') q(\mathbf{n}', \mathbf{n}), \ \text{for } \mathbf{n} \in S. \tag{1.1}$$

Equilibrium is achieved when the transient queue length distribution $\pi(\mathbf{n}, t)$ converges to a time independent form $\pi(\mathbf{n})$. The GBEs describe equilibrium in a network to occur when, for every possible state that the network can achieve, the probability flux out of the state due to customers entering and leaving nodes is balanced by the probability flux into that state due to customers entering and leaving nodes.

The equilibrium distribution of a network may be found by solving its GBEs, requiring the solution of a set of equations, with one equation for each state in the state space. In many circumstances the state space of a network is infinite, but even if it is constrained, it is usually extremely large, prohibiting direct computation of the equilibrium distribution from the GBEs. It is advantageous then to be able to model systems using queueing networks which have equilibrium distributions that are easy to compute. An example of such a class is *product form queueing networks.*

Almost all queue length equilibrium distributions for product form networks (at least all of the ones mentioned in this thesis) are of the form

$$\pi(\mathbf{n}) = K\phi(\mathbf{n})\prod_{i=1}^{N} f_i(\mathbf{c}_i), \quad \forall \mathbf{n} \in S, \tag{1.2}$$

where K is the normalising constant which maintains stochastic dependence between the nodes, $\phi(\cdot)$ is a given arbitrary function defined over the entire state space. The equilibrium distribution is a product over the nodes of the network of terms which are a function of the configuration of each node - hence product form queueing networks. The simple nature of the product form equilibrium distribution ensures that, once the normalising constant K is known, performance measures of interest are easy to calculate without lengthy numerical manipulations over the entire state space. If the normalising constant is finite, it is given by

$$K = \left[\sum_{\mathbf{n}\in S} \phi(\mathbf{n})\prod_{i=1}^{N} f_i(\mathbf{c}_i)\right]^{-1}. \tag{1.3}$$

A number of algorithms exist in the literature for calculating the normalising constants of product form queueing networks. See, for example, Buzen [9] and Coleman, Henderson and Taylor [16], [17] and [18]. These algorithms are continually being improved, to the extent that some do not need to enumerate over the entire state space, as equation (1.3) would suggest, but rather they take advantage of the specific structure of the product form queueing network to reduce the number of states over which the enumeration is performed.

When modelling complex systems, which need not have a product form equilibrium distribution, we are able to find approximate performance measures using aggregation and disaggregation techniques. These techniques arise following an adaption of Norton's Theorem, from electrical circuit theory, to queueing networks. This allows the behaviour of a subsystem to be studied by aggregating all components of the system, other than the subsystem of interest, into a single component. The parameters which dictate the interaction of the subsystem of interest with the rest of the system may be chosen so that the behaviour of the subsystem is not altered by the aggregation. Chandy, Herzog and Woo [10] and [11] provide the queueing network analogue of Norton's Theorem, and use it to approximate the equilibrium distributions of queueing networks which do not have product form. Other authors have refined the aggregation technique, and utilised its reverse process, namely disaggregation, to solve other queueing networks. We refer the interested reader to Balsamo and Iazeolla [1] and [2], Ciardo and Trivedi [15], Coyle, Henderson and Taylor [20] and Miyazawa [57]. These methods of simplification decompose the complex model until it has been reduced to a network of sub-models, each of which has a simple equilibrium distribution. The easy analysis of a product form queueing network makes it an ideal candidate to be a sub-model. Surprisingly, some decompositions provide exact, rather than approximate results.

This thesis will introduce, discuss, and demonstrate the applications of, a new set of queueing networks with product form equilibrium distributions, namely networks with negative customers, triggers and signals. These networks are new additions to the available library of sub-models which may be used in decompositions.

1.2 Literature Overview

1.2.1 Characterisation of Product Form Queueing Networks

Initially queueing theory was restricted to studying transient and steady state behaviour of single queues. In particular, Erlang and others derived the equilibrium distribution for the $M|M|s$ queue. This queue behaves in the following way: Customers arrive to the queue in a Poisson arrival stream at mean rate λ and receive a negative exponentially distributed service time with mean $\frac{1}{\mu}$, from s servers operating under the first come first served (FCFS) queueing discipline. The queue will be stable if $\lambda < \mu$, in which case the queue length equilibrium distribution is

$$\pi(n) = \begin{cases} K\left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}, & n \leq s \\ K\left(\frac{\lambda}{\mu}\right)^n \frac{1}{s!s^{n-s}}, & n > s, \end{cases} \tag{1.4}$$

where, again, K is the normalising constant.

Burke [7] and Reich [62] prove that the output of a stable $M|M|s$ queue is a Poisson stream with mean rate equal to that of the arrival stream. Jackson [46] notes that the sum of independent Poisson streams is a Poisson stream with mean rate equal to the sum of the original rates and uses the results of [7] and [62] to derive one of the first results on product form queueing networks. The networks of [46] consist of N nodes, the i-th node operating as an $M|M|s_i$ queue. Customers arrive to the i-th node at mean rate λ_i requesting mean service time $\frac{1}{\mu_i}$. Having received a complete service at node i a customer transfers (immediately) to node j with probability p_{ij}. The outside of the network is often referred to as node 0, and so

$$p_{i0} = 1 - \sum_{j=1}^{N} p_{ij}$$

is the probability that a customer will leave the network after being served at node i.

The marginal equilibrium distribution for node i, $\pi_i(\cdot)$, is equivalent to the equilibrium distribution of the node if it were isolated and fed with an increased arrival rate, to account for customers that arrive to the node from within the network rather than only from outside the network. The increased arrival rate is

$$Y_i = \lambda_i + \sum_{j=1}^{N} Y_j p_{ji}. \tag{1.5}$$

Thus $\pi_i(n_i)$ is given by equation 1.4 with Y_i replacing λ, service rate μ_i, and normalising constant K_i.

The main result of [46] is that the equilibrium distribution for these networks (given by equation (1.6)) is a product, over the nodes of the network, of the marginal equilibrium queue length distributions.

$$\pi(\mathbf{n}) = \prod_{i=1}^{N} \pi_i(n_i). \tag{1.6}$$

Thus equation (1.6) has the same form as equation (1.2) with

$$K = \prod_{i=1}^{N} K_i \text{ and } \phi(\mathbf{n}) = 1 \ \forall \mathbf{n} \in S.$$

Equations (1.5) are the traffic equations for Jackson networks. They state that the total arrival rate of customers to node i is the sum of the arrival rate of customers to node i from outside the network plus the rate at which customers are transferred to node i having received service somewhere in the network. In an open network, ($\lambda_i > 0$ for some i), Y_i is the throughput of customers at node i, but in a closed network it is the relative throughput. There is no distinction between total arrival rate and throughput in this instance, but we mention throughputs here because they play an important role in the interpretation of the traffic equations of other networks that will be presented in this thesis. Notice that equations (1.5) are linear in the Y_i's and are easily solved.

Jackson [47] generalises the previous network to allow the Poisson arrival rates to depend on the total number of customers in the system, and for the service rate at each node to depend arbitrarily on the number of customers at that node. These more general arrival rates allow, for example, the blocking and triggering of arrivals. This may be achieved by setting the total arrival rate $\lambda(n)$, when there are a total of n customers in the network, to be

$$\lambda(n) = \begin{cases} \infty, & n < n^-, \\ \lambda, & n^- \leq n < n^+, \\ 0, & n \geq n^+. \end{cases} \tag{1.7}$$

This would instantaneously trigger the arrival of a customer to the network if the number of customers present ever fell below n^-, and block customers from arriving if the network reaches its capacity of n^+. The service rate at node i is restricted to be positive except when there are no customers present: $\mu_i(n_i) > 0,\ n_i > 0,$ and $\mu_i(0) = 0$. The resulting queueing system, the so-called Jackson network, retains a product form equilibrium distribution.

Gordon and Newell [32] demonstrate the stochastic equivalence of a closed network of queues, with probabilistic routeing and negative exponential service times, to the networks of Jackson [46], so that the equilibrium distribution and traffic equations of [32] are given by equations (1.6) and (1.5), with $\lambda_i = 0,\ \forall i$.

Each of the queueing network papers mentioned thus far utilise proofs in which the GBEs are stated and the equilibrium distribution is obtained directly from them. In deriving the equilibrium distributions for certain nonlinear migration processes Whittle [74] and [75] introduces partial balance equations (PBEs), which are easier to solve than GBEs, and often provide a greater understanding of the system being studied. Whittle ensures that the GBE for a particular state is the sum of independent PBEs, one for each node in the network. These partial balance equations are sufficient (but not necessary) conditions for global balance. A solution to the partial balance equations is automatically a solution to the global balance equations, but it is sometimes the case, for non-Jackson networks, that the partial balance equations are inconsistent and therefore have no solution. The PBEs used in [75], which describes a closed network of queues, require that in each state the probability flux out of that state due to the departure of customers from a particular node be balanced by the probability flux into that state due to the arrival of customers to that same node. Whittle [74] extends [75] to provide a partial balance derivation for the product form equilibrium distribution of the equivalent open network of queues.

Many special cases of the queueing networks of [32], [46], [47], [74] and [75] appear in the literature, motivated by the concept of a computer system being a network of processors (CPU's, I/O devices) and a collection of customers (jobs, tasks). One of the most significant post-Jackson contributions to the literature on product form queueing networks is [3] by Baskett, Chandy, Muntz and Palacios in which it is pointed out that the queueing networks of Jackson [46], Buzen [8] and Gordon and Newell [32] are limited in their ability to model computer systems because

- all the customers are identical - restricting each unit in the model to follow the same behavioural rules,
- all of the service times are negative exponentially distributed.

Baskett *et al.* combine a number of previous disjoint results to obtain the so-called BCMP networks, which allow a finite number of distinct customer classes. The route-ing rule for a BCMP network is class dependent, and customers are allowed to change class as they route through the network. The Markov chain defined by the routeing rule is assumed to be decomposable into m ergodic sub-chains. Two types of state dependent arrival processes are allowed in an open BCMP network. Either the arrival process is Poisson, with mean arrival rate dependent on the total number of customers in the network, or each ergodic subchain has its own Poisson arrival process, with mean rate dependent on the number of customers in the subchain. In either case, arrivals will enter a particular node as a particular class of customer with a fixed probability. This is not the limit to the generality of BCMP networks. Each node in a BCMP network may operate according to one of four possible service disciplines:

- FCFS with all customer service times taken from the same negative exponential distribution, the rate of which may depend on the number of customers at the node,

- single server with processor sharing (with n customers present each customer will receive $\frac{1}{n}$ of the service effort), and each class of customer may have a distinct service time distribution,
- the number of customers at a node are restricted to be less than or equal to the number of servers, and each class of customer may have a distinct service time distribution,
- single server with pre-emptive resume last come first served (LCFS) scheduling, and each class of customer may have a distinct service time distribution.

It is also assumed in [3] that the service distributions of each node have a rational Laplace transform, but this assumption is unnecessary. The equilibrium distribution for BCMP networks is given by equation (1.2), where $\mathbf{c}_i$ is now a vector which details the number of customers of each class queued at node i. Baskett *et al.* use the method of partial balance to analyse the BCMP network, and, in particular, note that the GBEs are the sum of partial balance equations, which is generally the case for product form queueing networks.

Kelly [49] describes the need for different types of customers in many modelling situations. In particular, customers should be routed through the network according to their type, so that different behaviours may be modelled in a single system. The state of node i, when there are n_i customers present, is $\mathbf{x}_i = (t_i(1), t_i(2), \ldots, t_i(n_i))$, where $t_i(k)$ is the type of the customer at position k at node i. Kelly's model has the following features:

- customers of different types arrive to the network in independent Poisson streams,
- a total service effort of $\mu_i(n_i)$ is provided at node i,
- the proportion of service effort that the customer in position k at node i receives is $\gamma_i(k, \mathbf{x}_i)$,

- an arriving customer, which puts node i into state $\mathbf{x}_i$, moves directly to position k with probability $\delta_i(k, \mathbf{x}_i)$,
- there is one customer type for each distinct route through the network, so that a customer's type defines explicitly its destination following a service completion, and customers do not change type as they move through the network.

Using reversibility arguments, which will be described in Chapter 4 (see also Kelly [52]), Kelly [49] and [51] proves that this network has a product form equilibrium distribution if a set of traffic equations is satisfied, and if one of the following is true for each node in the network.

1. The service requirement of each customer at the node is taken from a negative exponential distribution with unit mean, and the parameters γ and δ are type independent. In this case $\mathbf{x}_i$ is replaced by n_i in $\gamma_i(k, \mathbf{x}_i)$ and $\delta_i(k, \mathbf{x}_i)$.
2. The service requirement of each customer at the node is taken from a negative exponential distribution with mean which may be type dependent, and $\gamma_i(k, \mathbf{x}_i) = \delta_i(k, \mathbf{x}_i),\ \forall k, \mathbf{x}_i$.

In the case of (2) the node is said to be symmetric in that, for any given state, the proportion of service effort given to a particular position in the queue is equal to the probability that an arriving customer is placed at that position. The network may consist of a mixture of nodes, satisfying (1) or (2), and retain product form.

Chandy, Howard and Towsley [12] characterise a set of product form systems in terms of station balance, requiring that the rate at which customers receive service at each position of a queue be proportional to the probability that a customer arrives at that position. This is precisely the formulation of Kelly's symmetric nodes. Chandy and Martin [13] extend the work of [12] by providing a characterisation of queueing networks in which the service disciplines are allowed to depend on the class as well as the position of the customer, as in the networks of Kelly [51]. Both [12] and

[13] do, however, generalise the work of Kelly by allowing the total service effort at a node to depend on $\mathbf{x}_i$ in the case of [12], and $\mathbf{n}$ in the case of [13], rather than just n_i, and by allowing customers to change types as they route through the network. Many other authors have introduced varying degrees of state dependence in the arrival rates, service disciplines and routeing probabilities to queueing networks and retained a product form equilibrium distribution. We will discuss their contributions to the literature in Chapter 3.

Systems with generally distributed lifetimes, such as BCMP networks, are often modelled as generalised semi-Markov processes (GSMP). A series of papers by Schassberger, [65], [66], [67] and [68], addresses the issue of insensitivity in GSMP to determine under what conditions networks of queues with generally distributed service lifetimes have product form. Schassberger provides a characterisation of discrete time product form processes in [69].

Muntz [58] examines the *$M \Rightarrow M$ property*, which is that Poisson arrivals imply Poisson departures. He shows that queueing networks which have this property have product form solutions, thus providing one of the first characterisations of product form queueing networks. The $M \Rightarrow M$ property concentrates on the total input and output of a system rather than its internal mechanics, and queueing networks with this property are called quasi-reversible. We provide a definition of quasi-reversibility in Chapter 4. The nodes of the so-called BCMP networks are quasi-reversible, as are those presented by Kelly in [49], [51] and [52]. The link between quasi-reversibility and product form queueing networks is that quasi-reversible nodes may be coupled together to form a process, which is itself quasi-reversible, and which has a product form equilibrium distribution. A proof of this appears in Kelly [53]. Networks of quasi-reversible nodes have been studied by a number of other researchers, mentioned briefly in Chapter 4. We note here that quasi-reversibility, partial balance and insensitivity are all characteristics that can be associated with product form queueing networks.

1.2.2 Signals: Negative Customers, Triggers and Interruptions

In the context of queueing networks, we define a signal to be an entity which causes a customer to leave the node at which it is queued before its service lifetime has expired. Signals may arrive from outside the network, or be routed through the network following a service completion. In the latter case, the customer, whose service lifetime has just expired, is transformed into a signal and is routed through the network. A number of papers have appeared in the literature on product form queueing networks which include entities that satisfy our definition of signals. These papers describe networks which include entities such as negative customers, triggers and interruptions.

Gelenbe [27] describes a queueing network with positive and negative customers. A negative customer arriving to a non-empty node will immediately force a normal (positive) customer to leave the network causing the queue length to decrease by one, but has no effect at an empty node. Gelenbe [27] allows negative customers to arrive to, be generated in, and routed around, a single server Jackson network with constant arrival and service rates. Positive and negative customers arrive to node i from outside the network in independent Poisson streams, at rates $\Lambda(i)$ and $\lambda(i)$ respectively. Positive customers receive a negative-exponentially distributed service time at rate $r(i)$. At the completion of its service, a positive customer may become a negative customer and be routed through the network. The routeing chain is type dependent, so that negative customers may be routed through the network in a different manner to positive customers. The routeing parameters of positive and negative customers are $p^+(i,j)$ and $p^-(i,j)$ respectively. Put simply a positive and a negative customer will eliminate each other if they meet. Gelenbe [27] does not allow negative customers to build up at a node, and so if they arrive to an empty node then they are lost from the network.

Gelenbe [27] writes that the original study [27] of queueing networks with positive and negative customers was motivated by an analogy of the network to a neural network.

> "In this analogy each queue represents a neuron. Positive customers moving from one queue to another represent excitation signals while negative customers going from one queue to another represent inhibition signals."

The queue length at each node represents the excitation level or potential of the neuron. Positive and negative customers are allowed to route between nodes, in the same way as adjacent neurons send and receive excitation and inhibition signals. The queue lengths at each node in the neural network model do not necessarily have 0 as a lower bound because negative neuron potentials may be required. For more detail on the application of these networks to neural networks see Gelenbe [25] and [26] and Kandel and Schwartz [48].

Gelenbe [27] shows that this network has a geometric product form equilibrium distribution of the form

$$\pi(\mathbf{n}) = \prod_{i=1}^{N}(1-q_i)q_i^{n_i}, \text{ where } q_i = \frac{\lambda^+(i)}{r(i)+\lambda^-(i)}, \tag{1.8}$$

λ_i^+ and λ_i^- are the total arrival rates of positive and negative customers to node i respectively, and the parameters of the network satisfy the traffic equations, similar to equations (1.5),

$$\lambda^+(i) = \sum_{j=1}^{N} q_j r(j) p^+(j,i) + \Lambda(i), \quad \lambda^-(i) = \sum_{j=1}^{N} q_j r(j) p^-(j,i) + \lambda(i). \tag{1.9}$$

This network will be discussed in detail in Chapter 2.

Gelenbe and Schassberger [31] prove the existence of a solution to equations (1.9), which is non-trivial since the equations are non-linear. They also provide conditions under which the resulting invariant measure provides a stationary queue length distribution for the network. Previously the existence and uniqueness of solution had only been established for feed-forward or hyper-stable networks. The networks of [31] also allow for multiple customer classes.

Gelenbe, Glynn and Sigman [28] describe the equilibrium distribution of, and give stability conditions for, a single server FCFS queue in which the superposition of positive and negative customer arrivals form a renewal process under two types of negative customer behaviour, namely the removal of the customer in service, and the removal of the customer at the tail of the queue. This corresponds to introducing negative customers to a standard FCFS single server queue with generally distributed inter-arrival and service times (GI/GI/1). The precise manner in which a negative customer eliminates a positive customer from a FCFS GI/GI/1 queue plays an important role in determining its stability. For example a node in which negative customers always remove the positive customer at the tail of the queue would be expected to be more stable then a node with the same parameters but in which the customer in service is removed because, on average, the server at the first node would be required to provide less service, per arrival to their queue, than the server in the second node. Some of the service effort of the server at the second node is wasted whenever a customer, which is in service, is eliminated.

Kelly [53] describes the transitions of a quasi-reversible process in terms of "arrivals" and "departures", explaining that these entities need not necessarily refer to normal (positive) customers. Although Kelly does not explicitly state it, a special case of a quasi-reversible node could occur when an "arrival" causes the queue length to decrease by one. In this instance the "arriving" entity is not a normal customer, but instead it behaves in the same manner as an arriving negative customer would. The theory of quasi-reversibility allows us to couple N of these nodes to form a quasi-reversible network, with a product form equilibrium distribution, similar to the networks of Gelenbe [27]. Further applications of quasi-reversibility to networks of queues with negative customers are given in Chapter 4.

Other networks have appeared in the literature in which customers leave a node prior to service completion. In particular, the notion that customers in a queueing network can be removed by exterior influences appears in Henderson and Taylor [33], in which networks of queues with interruptions were considered. In these networks, customers in service at a given node can have their service times ended prematurely by "interruptions", which arrive in Poisson streams from outside of the network, if the node satisfies one of the following criteria:

1. The node is symmetric, in which case the service times may be generally distributed.

2. The service times are negative-exponentially distributed.

The interruptions can be of different classes and the interrupted customer is routed through the network as a function of the class of interruption, the class of customer, and the queue at which the interruption occurred. One interpretation of events in such an interruption process is that they could be messages whose purpose is to cancel service and re-route customers along a different path from that which they would have followed had their service been allowed to finish naturally. Henderson and Taylor establish that such networks have product form equilibrium distributions.

Gelenbe [29] generalises the queueing networks of [27] by introducing triggers. A trigger arriving to a non-empty queue will force the immediate release (not elimination) of a positive customer to be routed through the network. The routeing chain that a customer follows after a triggered release, given by the parameters q_{ij}, may be different to the routeing chain followed by a customer that has received a complete service. The networks in [27] are a special case of those in [29] with all customers routed to the outside of the network following a triggered release. Triggers are more analogous to the interruptions of [33] than negative customers, in that they allow the triggered customer to be routed elsewhere in the network, but a positive customer eliminated by a negative customer must leave the network. The equilibrium distribution of a queueing network with triggered customer movement has the same form as that of a network with positive and negative customers, given by equation (1.8), but the traffic equations (and the GBEs) are different, and λ_i^- is now the total arrival rate of triggers to node i.

$$\lambda^+(i) = \sum_{j=1}^{N} q_j \left(r(j)p^+(j,i) + \lambda^-(j)q(j,i) \right) + \Lambda(i), \;\; \lambda^-(i) = \sum_{j=1}^{N} q_j r(j) p^-(j,i) + \lambda(i). \tag{1.10}$$

Further generalisations of [27] and [28] by other researchers will be mentioned later in this thesis.

The networks in [29] and [33] overlap somewhat. The customers in [33] may have generally distributed service times, but interruptions may only arrive from outside the network. The networks in [29] allow events within the network to interrupt customers elsewhere, but are limited to Poisson arrival streams and single server queues with negative exponential service times, and do not include multiple customer classes. We note that the so called G-networks of Gelenbe and Schassberger [31] include multiple customer classes, which would indicate that the networks of [29] may also be generalised to include them.

For further discussion on product form results for open and closed queueing networks with single customer movement we refer the interested reader to Kelly [52] and Walrand [73].

1.3 Outline

A large number of papers based on queueing networks with negative customers have appeared in the literature since we began working on this thesis. Chapter 2 is largely based on Henderson, Northcote and Taylor [40]. It is a direct generalisation on the work of Gelenbe in [27] and allows negative customers to arrive at a node in batches of any size. Gelenbe also realised the possibility for this generalisation and concurrently derived a very similar result [30] to that presented in Chapter 2. The differences between the networks of [30] and [40] are highlighted in that chapter.

In discussing insensitivity and the product form decomposability of processes with interruptions, Miyazawa [57] states that Gelenbe's model of a single movement queueing network with positive and negative customers can be extended to networks with multiple customer classes. This is achieved in Henderson [41], which is a generalisation of [30] and [40] to include state dependent service rates and the possibility for negative customers to accumulate at nodes. Networks with positive and negative customers and multiple customer classes have also been studied by Forneau and Gelenbe [23] and Gelenbe and Schassberger [31]. With the large number of papers emerging on queueing networks with some form of interactive interruption (signalling) the terminology and notation has become extremely confusing. Chapter 3, based on Henderson, Northcote and Taylor [42], combines the results of [27], [29], [30], [40] and [41] to describe a queueing network with signals and state dependent service rates. Here signals may be negative customers (single or in batches) or triggers. Some applications of queueing networks with signals, to areas as diverse as manufacturing processes, biological systems and communications protocols, are given in Chapter 3.

Quasi-reversibility is a common characteristic of most product form queueing networks. In Chapter 4 we present some results of Henderson, Pearce, Schassberger and Taylor [44] which show that the queueing networks with negative customers of Chapter 2 are quasi-reversible. We extend previous results on the coupling of quasi-reversible nodes to show that the queueing networks with signals, from Chapter 3, are also quasi-reversible.

The culmination of the work towards this thesis is presented in Chapter 5, based on Henderson, Northcote and Taylor [43], which applies the signalling concepts to batch movement queueing networks. Chapter 5 includes an application in which the queueing network results lead to a new class of stochastic Petri nets which have product form. A general framework for resource allocation problems is presented and is used to model a circuit switched network.

Chapter 2

Batch Destruction of Customers

Gelenbe [27] introduces negative customers to the literature on product form queueing networks by applying them to Jackson networks of single server queues, ($s_i = 1,\ \forall i$), with state independent arrival and service rates. Whereas a normal (or positive) customer increases the queue length of any queue it joins by one, a negative customer arriving to a non-empty queue will reduce the length of that queue by one customer, but has no effect when arriving at a queue which is already empty.

As mentioned earlier, the initial motivation for creating networks with positive and negative customers was in the modelling of neural networks. Applications for networks with positive and negative customers can be found in many other fields. In resource allocation problems an arrival of a positive customer could represent a request for a set of resources, and an arrival of a negative customer could represent a decision to cancel such a request. Several communications protocols may also be modelled using queueing networks incorporating negative customers. For example consider the "Go back N" protocol for data packet transmission. When transmitting messages between buffers using this protocol, the source buffer stores N data packets until it receives an "all clear" signal indicating that all packets have arrived intact at the destination buffer. To model the number of data packets stored in a network of communicating buffers, positive customer arrivals correspond to the arrival of data packets to a buffer, and negative customer arrivals correspond to an "all clear" signal deleting N customers from the buffer. This last application requires negative customers to be routed through the network in batches. This is not possible in the networks of [27].

We initially became interested in queueing networks with negative customers when we attempted to summarise the way in which the boundaries of the state space in the network of [27] impose partial balance. From our initial investigation we conjectured that more than one positive customer could be eliminated by negative customers in a single transition. We wrote down the partial balance equations that would be enforced by the boundaries of the state space (we allowed only non-negative queue lengths) for a network in which a customer moving through the network could cause the queue length at its destination node to decrease instantaneously by 2. This may seem to be a trivial generalisation, but the partial balance equations were non-linear and we were uncertain as to whether or nor it would be possible to determine a unique invariant measure from them that would satisfy the GBE. The results of numerical investigation were promising and this chapter presents the theory that proves our initial conjecture.

We generalise the Gelenbe [27] network of queues by allowing any customer moving through the network to be transformed instantaneously into a batch of negative customers upon arrival at its destination node. The batch size is determined by an arbitrary probability distribution. Customers may also be routed through the network as positive (normal) customers. We give a condition for this more general network (consisting of N nodes labelled from a set $\mathcal{L} = \{1, 2, \ldots, N\}$) to have a geometric product form equilibrium distribution. When this happens the N parameters of the equilibrium distribution satisfy a set of 2^N equations which are imposed by the boundaries of the state space (the hyper-planes $\{n_i = 0$, for $i \in \mathcal{E}(\mathbf{n})$, $\mathcal{E}(\mathbf{n}) \subseteq \mathcal{L}\}$, where n_i = number of customers at node i). These equations are shown to be equivalent to a set of N traffic equations representing expressions for the throughputs of the N servers in the network. Thus the parameters of the geometric distribution can be identified as the throughputs of the nodes. Furthermore it is established that the global balance equations for the process can always be written in terms of $(N + 1)$ partial balance equations which are individually satisfied by the equilibrium distribution of the network.

It has come to our attention, since the original submission of [40], the paper on which this chapter is based, that Gelenbe [30] also shows that networks of queues with batch routeing of negative customers have a geometric product form equilibrium distribution. Gelenbe's model is a little less general then the one that we present here, in that it assumes that an arriving customer removes a batch of customers from queue i with a probability distribution that does not depend on its node of origin. In contrast our model, as described in Section 2.1, allows the distribution of the number of customers removed to depend on both the origin and destination node of the moving customer.

2.1 System Description

The model consists of a network of N *single* server nodes, indexed by a set of node labels $\mathcal{L} = \{1, 2, \ldots, N\}$, with the outside of the network referred to as node 0 for notational simplicity. Positive customers at node i request a negative exponentially distributed service time with mean $\frac{1}{\mu_i}$. i.e. positive customers are emitted from node i, if any are present, at rate μ_i, $i \in \mathcal{L} \cup \{0\}$. Once emitted from node i, (including node 0), a positive customer will, with probability

- d^+_{ij} be transferred to node $j \in \mathcal{L} \cup \{0\}$ (increasing the queue length by if $j \in \mathcal{L}$)
- $d^-_{ij}(k)$, $k \geq 1$ be eliminated from the network and simultaneously signal the arrival of a batch of k negative customers to node $j \in \mathcal{L}$.

The latter case may be thought of as a request for the elimination of k customers from node j, due to a service completion at node i. In response to such a request k customers will be eliminated from node j, or, if there are less than k customers at node j, then node j will be emptied. These probabilities also apply to emissions from node 0, and so an outside emission, (an arrival to the network), will signal the arrival of a batch of k negative customers to node i with probability $d_{0i}^{-}(k)$, and will increase the queue length at node i by one with probability d_{0i}^{+}. The parameters d_{ij}^{+} and $d_{ij}^{-}(k)$ of the system are probabilities and so

$$\sum_{j=0}^{N} d_{ij}^{+} + \sum_{j=1}^{N} \sum_{k=1}^{\infty} d_{ij}^{-}(k) = 1, \ \forall i \in \mathcal{L} \cup \{0\}. \tag{2.1}$$

Consider the ways in which a single customer leaves the queue at node i and departs the network without affecting the queue lengths of any other node in the network. This may occur in any of the following ways:

- an emission from node 0 signals the arrival of one negative customer to node i (with rate $\mu_0 d_{0i}^{-}(1)$),
- following a service completion at node i a customer is routed to the outside of the network, either as a positive customer, or a batch of negative customers,
- an emission from node i signals the arrival of a batch of negative customers to node j when the queue at node j is empty.

Thus many events have the same net effect on the network. To avoid some duplicity of intensities, and for simplicity of notation, we assume that

$$d_{ii}^{+} = d_{i0}^{-}(k) = 0, \ \forall i \in \mathcal{L} \cup \{0\}, \ k \geq 1. \tag{2.2}$$

We have found that the equations governing the process are more compact with this notation. We also assume that $\mu_j d_{ji}^{+} > 0$ for at least one $j \in \mathcal{L}$ for each node i, so that it is possible for the number of customers at node i to increase by 1. This ensures the irreducibility of the state space.

2.2 The Equilibrium Distribution

At time t let the queue length at node i be $n_i(t)$. We form a continuous time Markov chain $\{\mathbf{n}(t) = (n_1(t), n_2(t), \cdots, n_N(t)) : t \geq 0\}$, which satisfies the usual Chapman-

Kolmogorov equations. We look for an equilibrium invariant measure of the form

$$\pi(\mathbf{n}) = K \prod_{i=1}^{N} \left(\frac{Y_i}{\mu_i}\right)^{n_i}. \tag{2.3}$$

K can be chosen so that this sums to one if and only if $Y_i < \mu_i, \ \forall i \in \mathcal{L}$. In this case we show that Y_i can be interpreted as the throughput of the single server at node i, or equivalently the portion of the arrival stream that actually receives a complete service from the server. We set $Y_0 = \mu_0$ so that equation (2.3) may be written as a product over $N+1$ nodes, from 0 to N, and the state of the outside of the network (node 0) will not affect the invariant measure.

Let $\mathbf{e}_i$ be a $N \times 1$ vector of zeros except for a 1 in the i-th position, and $\mathbf{e}_0$ be the corresponding zero vector. Consider the state space $S = \{\mathbf{n} : n_i \geq 0, \ \forall i \in \mathcal{L}\}$. For each $\mathbf{n} \in S$, let $\mathcal{E}(\mathbf{n}) \subseteq \mathcal{L}$ be the set of nodes which are empty in state $\mathbf{n}$ and $\mathcal{E}(\mathbf{n})^c = \mathcal{L} \backslash \mathcal{E}(\mathbf{n})$ be the set of non empty nodes, so that $n_i = 0$ if $i \in \mathcal{E}(\mathbf{n})$, and $n_i > 0$ if $i \in \mathcal{E}(\mathbf{n})^c$. To leave state $\mathbf{n}$ one of the following transitions must occur

- an emission from node 0 arrives at a non-empty node, and has a net effect on the number of customers at that node at rate

$$\mu_0 \sum_{i \in \mathcal{E}(\mathbf{n})^c} \left(d_{0i}^+ + \sum_{\ell=1}^{\infty} d_{0i}^-(\ell) \right)$$

- an emission from node 0 deposits a positive customer at an empty node at rate

$$\mu_0 \sum_{i \in \mathcal{E}(\mathbf{n})} d_{0i}^+$$

- an emission occurs at a non-empty node at rate

$$\sum_{i \in \mathcal{E}(\mathbf{n})^c} \mu_i$$

Transitions into state $\mathbf{n}$ occur from all states of the form

- $\mathbf{n} + \mathbf{e}_h$, $h \in \mathcal{L}$ due to an emission from node h leaving the network or being transferred as a batch of negative customers to an empty node at rate

$$\mu_h \left(d_{h0}^+ + \sum_{i \in \mathcal{E}(\mathbf{n})} \sum_{\ell=1}^{\infty} d_{hi}^-(\ell) \right)$$

- $\mathbf{n} + \mathbf{e}_h - \mathbf{e}_i$, $h \in \mathcal{L} \cup \{0\}$, $i \in \mathcal{E}(\mathbf{n})^c$, due to an emission from node h depositing a positive customer at node i at rate

$$\mu_h d_{hi}^+$$

- $\mathbf{n} + \mathbf{e}_h + \ell \mathbf{e}_i$, $h \in \mathcal{L} \cup \{0\}$, $i \in \mathcal{E}(\mathbf{n})^c$, $\ell \geq 1$, due to an emission from node h depositing a batch of ℓ negative customers at node i at rate

$$\mu_h d_{hi}^-(\ell)$$

- $\mathbf{n} + \mathbf{e}_h + \ell \mathbf{e}_i$, $h \in \mathcal{L} \cup \{0\}$, $i \in \mathcal{E}(\mathbf{n})$, $\ell \geq 1$, due to an emission from node h depositing a batch of $k \geq \ell$ negative customers at node i at rate

$$\mu_h \sum_{k=\ell}^{\infty} d_{hi}^-(k)$$

Thus the global balance equations relating flux into and out of state $\mathbf{n}$ are

$$\begin{aligned}
\pi(\mathbf{n}) & \left[\sum_{i \in \mathcal{E}(\mathbf{n})^c} \left(\mu_0 \left(d_{0i}^+ + \sum_{\ell=1}^{\infty} d_{0i}^-(\ell) \right) + \mu_i \right) + \sum_{i \in \mathcal{E}(\mathbf{n})} \mu_0 d_{0i}^+ \right] \\
&= \sum_{i=1}^{N} \pi(\mathbf{n} + \mathbf{e}_i) \mu_i \left(d_{i0}^+ + \sum_{h \in \mathcal{E}(\mathbf{n})} \sum_{\ell=1}^{\infty} d_{ih}^-(\ell) \right) \\
&\quad + \sum_{h=0}^{N} \mu_h \left[\sum_{i \in \mathcal{E}(\mathbf{n})^c} \left(\pi(\mathbf{n} + \mathbf{e}_h - \mathbf{e}_i) d_{hi}^+ + \sum_{\ell=1}^{\infty} \pi(\mathbf{n} + \mathbf{e}_h + \ell \mathbf{e}_i) d_{hi}^-(\ell) \right) \right. \\
&\quad \left. + \sum_{i \in \mathcal{E}(\mathbf{n})} \sum_{\ell=1}^{\infty} \pi(\mathbf{n} + \mathbf{e}_h + \ell \mathbf{e}_i) \sum_{k=\ell}^{\infty} d_{hi}^-(k) \right]. \qquad (2.4)
\end{aligned}$$

Substituting from equation (2.3) and simplifying (dividing by $\pi(\mathbf{n})$) gives

$$\sum_{i \in \mathcal{E}(\mathbf{n})^c} \left(\mu_0 \left(d_{0i}^+ + \sum_{\ell=1}^{\infty} d_{0i}^-(\ell) \right) + \mu_i \right) + \sum_{i \in \mathcal{E}(\mathbf{n})} \mu_0 d_{0i}^+$$

$$= \sum_{i=1}^{N} Y_i \left(d_{i0}^{+} + \sum_{h\in\mathcal{E}(\mathbf{n})} \sum_{\ell=1}^{\infty} d_{ih}^{-}(\ell) \right) + \sum_{h=0}^{N} Y_h \left[\sum_{i\in\mathcal{E}(\mathbf{n})^c} \left(\left(\tfrac{\mu_i}{Y_i}\right) d_{hi}^{+} + \sum_{\ell=1}^{\infty} \left(\tfrac{Y_i}{\mu_i}\right)^{\ell} d_{hi}^{-}(\ell) \right) + \sum_{i\in\mathcal{E}(\mathbf{n})} \sum_{\ell=1}^{\infty} \left(\tfrac{Y_i}{\mu_i}\right)^{\ell} \sum_{k=\ell}^{\infty} d_{hi}^{-}(k) \right]. \quad (2.5)$$

That is, if the equilibrium distribution is to be given by equation (2.3) then $\{Y_i,\ i \in \mathcal{L}\}$ must satisfy equation (2.5). We refer to equation (2.5) as the equation imposed by the boundary of the state space, given by the hyper-plane $\{\mathbf{n} : n_i = 0\ i \in \mathcal{E}(\mathbf{n}),\ n_i > 0\ i \in \mathcal{E}(\mathbf{n})^c\}$. Note that equation (2.5) depends upon the state of the network only through those nodes which are empty and in its present form represents 2^N equations, one for each distinct $\mathcal{E}(\mathbf{n}) \subseteq \mathcal{L}$. In the following we show that these 2^N equations can be replaced by N state independent equations, one for each node in the network. Rearranging equation (2.5), noting that $\sum_{i=1}^{N} = \sum_{i\in\mathcal{E}(\mathbf{n})} + \sum_{i\in\mathcal{E}(\mathbf{n})^c}$, we get

$$\sum_{i\in\mathcal{E}(\mathbf{n})^c} A_i = \sum_{i\in\mathcal{E}(\mathbf{n})} B_i, \text{ where}$$

$$A_i = \mu_0 \left(d_{0i}^{+} + \sum_{\ell=1}^{\infty} d_{0i}^{-}(\ell) \right) + \mu_i - Y_i d_{i0}^{+} - \sum_{h=0}^{N} Y_h \left(\left(\tfrac{\mu_i}{Y_i}\right) d_{hi}^{+} + \sum_{\ell=1}^{\infty} \left(\tfrac{Y_i}{\mu_i}\right)^{\ell} d_{hi}^{-}(\ell) \right),$$

$$B_i = -\mu_0 d_{0i}^{+} + Y_i d_{i0}^{+} + \sum_{h=1}^{N} Y_h \sum_{\ell=1}^{\infty} d_{hi}^{-}(\ell) + \sum_{h=0}^{N} Y_h \sum_{\ell=1}^{\infty} \left(\tfrac{Y_i}{\mu_i}\right)^{\ell} \sum_{k=\ell}^{\infty} d_{hi}^{-}(k).$$

Lemmas 2.2.1, 2.2.2 and 2.2.3 establish some basic relationships involving the expressions A_i and B_i. The results of these lemmas are utilised in the proof of Theorem 2.2.1, which is the main result of this chapter.

Lemma 2.2.1 *For all $i \in \mathcal{L}$,*

$$\left(\tfrac{Y_i}{\mu_i}\right)(A_i + B_i) = Y_i + \sum_{h=0}^{N} Y_h \sum_{\ell=1}^{\infty} \left(\tfrac{Y_i}{\mu_i}\right)^{\ell} \sum_{k=\ell}^{\infty} d_{hi}^{-}(k) - \sum_{h=0}^{N} Y_h d_{hi}^{+}.$$

Proof: Remembering that $Y_0 = \mu_0$,

$$\begin{aligned} \left(\tfrac{Y_i}{\mu_i}\right)(A_i + B_i) \;=\; & \left(\tfrac{Y_i}{\mu_i}\right) \mu_0 \sum_{\ell=1}^{\infty} d_{0i}^{-}(\ell) + Y_i - \sum_{h=0}^{N} Y_h \left(d_{hi}^{+} + \sum_{\ell=1}^{\infty} \left(\tfrac{Y_i}{\mu_i}\right)^{\ell+1} d_{hi}^{-}(\ell) \right) \\ & + \left(\tfrac{Y_i}{\mu_i}\right) \sum_{h=1}^{N} Y_h \sum_{\ell=1}^{\infty} d_{hi}^{-}(\ell) + \sum_{h=0}^{N} Y_h \sum_{\ell=1}^{\infty} \left(\tfrac{Y_i}{\mu_i}\right)^{\ell+1} \sum_{k=\ell}^{\infty} d_{hi}^{-}(k), \end{aligned}$$

$$
\begin{aligned}
&= Y_i + \left(\tfrac{Y_i}{\mu_i}\right) \sum_{h=0}^{N} Y_h \sum_{\ell=1}^{\infty} d_{hi}^{-}(\ell) + \sum_{h=0}^{N} Y_h \sum_{\ell=1}^{\infty} \left(\tfrac{Y_i}{\mu_i}\right)^{\ell+1} \sum_{k=\ell+1}^{\infty} d_{hi}^{-}(k) - \sum_{h=0}^{N} Y_h d_{hi}^{+}, \\
&= Y_i + \left(\tfrac{Y_i}{\mu_i}\right) \sum_{h=0}^{N} Y_h \sum_{\ell=1}^{\infty} d_{hi}^{-}(\ell) + \sum_{h=0}^{N} Y_h \sum_{\ell=2}^{\infty} \left(\tfrac{Y_i}{\mu_i}\right)^{\ell} \sum_{k=\ell}^{\infty} d_{hi}^{-}(k) - \sum_{h=0}^{N} Y_h d_{hi}^{+}, \\
&= Y_i + \sum_{h=0}^{N} Y_h \sum_{\ell=1}^{\infty} \left(\tfrac{Y_i}{\mu_i}\right)^{\ell} \sum_{k=\ell}^{\infty} d_{hi}^{-}(k) - \sum_{h=0}^{N} Y_h d_{hi}^{+}.
\end{aligned}
$$

□

Lemma 2.2.2

$$
\sum_{i=1}^{N} \left(\tfrac{Y_i}{\mu_i}\right)(A_i + B_i) = \sum_{i=1}^{N} B_i
$$

Proof: From Lemma 2.2.1, repeatedly utilising equation (2.1) and changing subscripts:

$$
\begin{aligned}
\sum_{i=1}^{N} \left(\tfrac{Y_i}{\mu_i}\right)(A_i + B_i) &= \sum_{i=1}^{N} Y_i + \sum_{i=1}^{N} \sum_{h=0}^{N} Y_h \sum_{\ell=1}^{\infty} \left(\tfrac{Y_i}{\mu_i}\right)^{\ell} \sum_{k=\ell}^{\infty} d_{hi}^{-}(k) - \sum_{i=1}^{N} \sum_{h=0}^{N} Y_h d_{hi}^{+}, \\
&= \sum_{h=0}^{N} Y_h \left(1 - \sum_{i=1}^{N} d_{hi}^{+}\right) - \mu_0 + \sum_{i=1}^{N} \sum_{h=0}^{N} Y_h \sum_{\ell=1}^{\infty} \left(\tfrac{Y_i}{\mu_i}\right)^{\ell} \sum_{k=\ell}^{\infty} d_{hi}^{-}(k), \\
&= \sum_{h=0}^{N} Y_h \left(d_{h0}^{+} + \sum_{i=1}^{N} \sum_{\ell=1}^{\infty} d_{hi}^{-}(\ell)\right) - \mu_0 + \sum_{i=1}^{N} \sum_{h=0}^{N} Y_h \sum_{\ell=1}^{\infty} \left(\tfrac{Y_i}{\mu_i}\right)^{\ell} \sum_{k=\ell}^{\infty} d_{hi}^{-}(k), \\
&= \sum_{i=1}^{N} \left[-\mu_0 d_{0i}^{+} + Y_i d_{i0}^{+} + \sum_{h=1}^{N} Y_h \sum_{\ell=1}^{\infty} d_{hi}^{-}(\ell) + \sum_{h=0}^{N} Y_h \sum_{\ell=1}^{\infty} \left(\tfrac{Y_i}{\mu_i}\right)^{\ell} \sum_{k=\ell}^{\infty} d_{hi}^{-}(k)\right], \\
&= \sum_{i=1}^{N} B_i.
\end{aligned}
$$

□

Lemma 2.2.3 *For all $\mathcal{E} \subseteq \mathcal{L}$, $\{Y_i,\ i \in \mathcal{L}\}$ satisfy*

$$
\sum_{j \in \mathcal{E}^c} A_j = \sum_{j \in \mathcal{E}} B_j, \tag{2.6}
$$

if and only if they also satisfy

$$
A_i + B_i = 0, \quad \forall i \in \mathcal{L}. \tag{2.7}
$$

Proof: Let $\mathcal{E}_1 = \mathcal{E}_2 \cup \{i\}$ for some $i \notin \mathcal{E}_2$. Assume equation (2.6) holds for all $\mathcal{E} \subseteq \mathcal{L}$. Then

$$\sum_{j \in \mathcal{E}_1^c} A_j = \sum_{j \in \mathcal{E}_1} B_j, \text{ and therefore}$$
$$\sum_{j \in \mathcal{E}_2^c} A_j - A_i = \sum_{j \in \mathcal{E}_2} B_j + B_i. \tag{2.8}$$

Since equation (2.6) holds for $\mathcal{E}_2$ we immediately obtain equation (2.7) for a particular i. We can repeat this for all $i \in \mathcal{L}$. That is, equation (2.6) implies equation (2.7). For the converse we use induction.

Assume that we have found $\{Y_i,\ i \in \mathcal{L}\}$ which satisfy equation (2.7) and also satisfy equation (2.6) for all sets $\mathcal{E} \subseteq \mathcal{L} : |\mathcal{E}| = k$ for some $k > 0$. Choose a particular $\mathcal{E}_2 : |\mathcal{E}_2| = k-1$ and consider equation (2.8) again. In particular observe that equation (2.7) holds for all $i \in \mathcal{L}$, and with $\mathcal{E}_1 = \mathcal{E}_2 \cup \{i\},\ i \notin \mathcal{E}_2$, we have assumed that equation (2.6) holds for $\mathcal{E}_1$. Then equation (2.8) combined with equation (2.7) guarantees that equation (2.6) must hold for any choice of $\mathcal{E}_2$ and i such that $|\mathcal{E}_2| = k-1$. Thus we have that equation (2.6) holds for all $\mathcal{E} \subset \mathcal{L} : |\mathcal{E}| \leq k$, establishing the induction argument and we now need only a starting point to complete the proof.

Consider the case $|\mathcal{E}| = N$, that is the set $\mathcal{E} = \mathcal{L} = \{1, 2, \ldots, N\}$. We need to show that equation (2.6) can be obtained from equations (2.7) for this particular choice of $\mathcal{E}$, that is we need to show that $\sum_{i=1}^N B_i = 0$. Assume equation (2.7) holds, then

$$\left(\frac{Y_i}{\mu_i}\right)(A_i + B_i) = 0.$$

Summing over i and using Lemma 2.2.2 we have

$$\sum_{i=1}^N \left(\frac{Y_i}{\mu_i}\right)(A_i + B_i) = \sum_{i=1}^N B_i = 0.$$

That is equation (2.6) holds for $\mathcal{E} = \mathcal{L} = \{1, 2, \ldots, N\}$, thus completing the induction.

□

Theorem 2.2.1 *For the queueing network described in Section 2.1, equation (2.3) is an invariant measure if and only if there exists a non-negative solution,* $\{Y_i,\ i \in \mathcal{L}\}$, *to the following non-linear system of equations*

$$Y_i = \sum_{h=0}^{N} Y_h d_{hi}^{+} - \sum_{h=0}^{N} Y_h \sum_{\ell=1}^{\infty} \left(\frac{Y_i}{\mu_i}\right)^{\ell} \sum_{k=\ell}^{\infty} d_{hi}^{-}(k), \; i \in \mathcal{L}. \tag{2.9}$$

Proof: Equations (2.5) and (2.6) are equivalent and Lemma 2.2.1 indicates that equation (2.9) is equivalent to $\left(\frac{Y_i}{\mu_i}\right)(A_i + B_i) = 0$. Thus Lemma 2.2.3 verifies that the N equations (2.9) can hold if and only if the 2^N equations (2.5) also hold. Thus $\{Y_i, \; i \in \mathcal{L}\}$ satisfy equations (2.5), which are the global balance equations when multiplied by $\pi(\mathbf{n})$ from equation (2.3).

□

Gelenbe [30] includes a proof for the existence of an invariant measure for a network of queues similar to the one presented in this chapter. The proof is, however, incomplete since he does not show that the equation

$$q_i = \frac{\lambda^+(i)}{r(i) + \lambda^-(i) f_i(q_i)},$$

where

$$f_i(x) = \frac{1 - \sum_{k=1}^{\infty} \pi_{ik} x^k}{1 - x},$$

can be solved explicitly for $\{q_i, \; i \in \mathcal{L}\}$ and so

$$\lambda^+(i) = \sum_{j=1}^{N} \lambda^+(j) g(j) p^+(j,i) + \Lambda(i), \quad \lambda^-(i) = \sum_{j=1}^{N} \lambda^+(j) g(j) p^-(j,i) + \lambda(i),$$

where

$$g(i) = \frac{r(i)}{r(i) + \lambda^-(i) f_i(q_i)},$$

cannot be regarded as equations only in terms of the variables $\{\lambda^+(i), \lambda^-(i), \; i \in \mathcal{L}\}$. Henderson, Northcote and Taylor [42] give a sufficient condition for the existence of an invariant measure, which could be adapted to the network discussed here, but do not prove existence unconditionally. We do so here, by extending the existence proof of Gelenbe [30].

Theorem 2.2.2 *There exists a unique non negative solution $\{Y_i,\ i \in \mathcal{L}\}$ to equations (2.9).*

Proof: It is known from the theory of Markov processes [64] that a regular irreducible Markov process has a unique invariant measure up to constant multiples. Therefore if there exists a solution $\{Y_i,\ i \in \mathcal{L}\}$ to equations (2.9) then it is unique. Rearrange equation (2.9) to obtain

$$Y_i = \lambda_i^+ - \sum_{\ell=1}^{\infty} \left(\tfrac{Y_i}{\mu_i}\right)^{\ell} \sum_{k=\ell}^{\infty} \lambda_i^-(k),\ i \in \mathcal{L}, \tag{2.10}$$

where

$$\lambda_i^+ = \sum_{h=0}^{N} Y_h d_{hi}^+,\ i \in \mathcal{L}, \tag{2.11}$$

and

$$\lambda_i^-(k) = \sum_{h=0}^{N} Y_h d_{hi}^-(k),\ i \in \mathcal{L},\ k \geq 1. \tag{2.12}$$

Rewrite equation (2.10) as a polynomial in Y_i.

$$F_i(Y_i) = Y_i - \lambda_i^+ + \sum_{\ell=1}^{\infty} \left(\tfrac{Y_i}{\mu_i}\right)^{\ell} \sum_{k=\ell}^{\infty} \lambda_i^-(k),\ i \in \mathcal{L}. \tag{2.13}$$

For each $i \in \mathcal{L}$, given fixed non-negative values of λ_i^+ and $\lambda_i^-(k),\ k \geq 1,\ Y_i$ may be determined uniquely from equation (2.13) as a root of the polynomial $F_i(\cdot)$. This may be shown by noting the following and applying the mean value theorem and intermediate value theorem,

$$F_i(0) \leq 0,$$

$$\lim_{t\to\infty} F_i(t) > 0,$$

$$F_i'(x) > 0,\ x \geq 0.$$

Now let $q_i = \left(\frac{Y_i}{\mu_i}\right),\ i \in \mathcal{L},\ q_0 = 1$, so that equations (2.10) become

$$\begin{aligned} q_i\mu_i &= \lambda_i^+ - \sum_{\ell=1}^{\infty} q_i^{\ell} \sum_{k=\ell}^{\infty} \lambda_i^-(k),\ i \in \mathcal{L}, \\ &= \lambda_i^+ - \sum_{k=1}^{\infty} \lambda_i^-(k) \sum_{\ell=1}^{k} q_i^{\ell},\ i \in \mathcal{L}. \end{aligned}$$

Thus

$$q_i = \frac{\lambda_i^+}{\mu_i + \sum_{k=1}^{\infty} \lambda_i^-(k) f_k(q_i)}, \ i \in \mathcal{L}, \tag{2.14}$$

where we use an alternative form to that of Gelenbe [30] for the functions f_k, namely

$$f_k(x) = \sum_{\ell=1}^{k} x^{\ell-1}, \ k \geq 1.$$

Let

$$\begin{aligned} g_i &= \frac{Y_i}{\lambda_i^+}, \ i \in \mathcal{L}, \\ &= \frac{\mu_i}{\mu_i + \sum_{k=1}^{\infty} \lambda_i^-(k) f_k(q_i)}, \ i \in \mathcal{L}. \end{aligned} \tag{2.15}$$

Thus, for each $i \in \mathcal{L}$, given fixed non-negative values of λ_i^+ and $\lambda_i^-(k), \ k \geq 1$, $q_i = \frac{Y_i}{\mu_i}$ and $g_i = \frac{Y_i}{\lambda_i^+}$ may be determined uniquely. Equations (2.11) and (2.12) become

$$\begin{aligned} \lambda_i^+ &= \sum_{h=1}^{N} \lambda_h^+ g_h d_{hi}^+ + \mu_0 d_{0i}^+, \ i \in \mathcal{L}, \\ \lambda_i^-(k) &= \sum_{h=1}^{N} \lambda_h^+ g_h d_{hi}^-(k) + \mu_0 d_{0i}^-(k), \ i \in \mathcal{L}, \ k \geq 1. \end{aligned}$$

These last equations may now be written in vector form as follows

$$\boldsymbol{\lambda}^+ = \boldsymbol{\lambda}^+ G D^+ + \boldsymbol{\mu}^+, \tag{2.16}$$

$$\boldsymbol{\lambda}^-(k) = \boldsymbol{\lambda}^+ G D_k^- + \boldsymbol{\mu}_k^-, \ k \geq 1, \tag{2.17}$$

where $\boldsymbol{\lambda}^+$, $\boldsymbol{\lambda}^-(k)$, $\boldsymbol{\mu}^+$ and $\boldsymbol{\mu}_k^-$ are $1 \times N$ row vectors whose i-th entries are λ_i^+, $\lambda_i^-(k)$, $\mu_0 d_{0i}^+$ and $\mu_0 d_{0i}^-(k)$ respectively and G, D^+ and D_k^- are $N \times N$ matrices given by

$$G = diag(g_i), \ D^+ = \left[d_{ij}^+\right]_{ij} \text{ and } D_k^- = \left[d_{ij}^-(k)\right]_{ij}.$$

Equations (2.17) may be combined by forming

$$\boldsymbol{\lambda}^- = \left[\boldsymbol{\lambda}^-(1), \boldsymbol{\lambda}^-(2), \ldots\right], \ D^- = \left[D_1^- \ D_2^- \ \ldots\right] \text{ and } \boldsymbol{\mu}^- = \left[\boldsymbol{\mu}_1^-, \boldsymbol{\mu}_2^-, \ldots\right],$$

so that

$$\boldsymbol{\lambda}^- = \boldsymbol{\lambda}^+ G D^- + \boldsymbol{\mu}^-. \tag{2.18}$$

From equation (2.16)

$$\boldsymbol{\lambda}^+(I - GD^+) = \boldsymbol{\mu}^+.$$

Noting that $G \leq I$ component-wise, since $g_i \leq 1,\ i \in \mathcal{L}$, by equation (2.10), and that D^+ is sub-stochastic without any ergodic classes, then $(I - GD^+)$ must be invertible. Therefore we can solve for $\boldsymbol{\lambda}^+$ and $\boldsymbol{\lambda}^-$ explicitly to get

$$\boldsymbol{\lambda}^+ = \boldsymbol{\mu}^+ \sum_{\ell=0}^{\infty} (GD^+)^\ell, \tag{2.19}$$

$$\boldsymbol{\lambda}^- = \boldsymbol{\mu}^+ \sum_{\ell=0}^{\infty} (GD^+)^\ell GD^- + \boldsymbol{\mu}^-. \tag{2.20}$$

We combine equations (2.19) and (2.20), by setting $\boldsymbol{\lambda} = \left[\boldsymbol{\lambda}^+, \boldsymbol{\lambda}^-\right]$ and $\boldsymbol{\mu} = [\mathbf{0}, \boldsymbol{\mu}^-]$, to obtain

$$\boldsymbol{\lambda} = \boldsymbol{\mu}^+ \sum_{\ell=0}^{\infty} (GD^+)^\ell \left[I \quad GD^-\right] + \boldsymbol{\mu}. \tag{2.21}$$

Let $\mathbf{x} = \boldsymbol{\lambda} - \boldsymbol{\mu}$ and define the function $F(\cdot)$ to be

$$F(\mathbf{x}) = \boldsymbol{\mu}^+ \sum_{\ell=0}^{\infty} (GD^+)^\ell \left[I \quad GD^-\right] + \boldsymbol{\mu}, \tag{2.22}$$

where $F(\cdot)$ depends on $\mathbf{x}$ through the matrix G. For non-negative vectors $\mathbf{x}$, $F(\mathbf{x})$ is a non-negative continuous function and is maximised, for all components, when $G = I$, since all other terms in equation (2.22) are non-negative constants. Let this maximum value of $F(\cdot)$, corresponding to $G = I$, be F^*, so that $F(\cdot)$ satisfies

$$F : [\mathbf{0}, F^*] \longrightarrow [\mathbf{0}, F^*].$$

Thus, by Brouwer's fixed point theorem [22], there exists a fixed point $\mathbf{x}^*$ satisfying $\mathbf{x}^* = F(\mathbf{x}^*)$. From this fixed point we can obtain fixed vectors $\boldsymbol{\lambda}^-$, $\boldsymbol{\lambda}^+$ and ultimately, from equation (2.15), $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N) = \boldsymbol{\lambda}^+ G$.

□

In the case where equation (2.3) is an invariant measure for the network, and $Y_i < \mu_i$ for all $i \in \mathcal{L}$, ensuring summability, we set

$$K = \prod_{i=1}^{N} \left(1 - \left(\tfrac{Y_i}{\mu_i}\right)\right)$$

in equation (2.3) making it the equilibrium distribution of the network and the following interpretations may be made

$$\left(\frac{Y_i}{\mu_i}\right)^{\ell} = \Pr(\text{node } i \text{ contains at least } \ell \text{ customers}),$$

$$\left(1-\left(\frac{Y_i}{\mu_i}\right)\right)\left(\frac{Y_i}{\mu_i}\right)^{\ell} = \Pr(\text{node } i \text{ contains exactly } \ell \text{ customers}).$$

Remark 2.2.1 *The parameter Y_i has an interpretation as the throughput of the single server at node i.*

Putting $Y_0 = \mu_0$, consider the balance required between the rate at which single positive customers enter node i and the rate at which customers leave node i, either due to a service completion, or after being eliminated by the arrival of a negative customer. This balance is given, for all $i \in \mathcal{L}$, by

$$\sum_{j=0}^{N} Y_j d_{ji}^{+} = Y_i + \sum_{j=0}^{N} Y_j \left[\sum_{\ell=1}^{\infty} \ell \left(\frac{Y_i}{\mu_i}\right)^{\ell+1} d_{ji}^{-}(\ell) + \sum_{\ell=1}^{\infty} \ell \left(1-\left(\frac{Y_i}{\mu_i}\right)\right)\left(\frac{Y_i}{\mu_i}\right)^{\ell} \sum_{k=\ell}^{\infty} d_{ji}^{-}(k)\right]. \quad (2.23)$$

The RHS of equation (2.23) consists of terms describing the rate of which customers leave node i. This can occur through service completions, (the Y_i term), or due to the effects of negative customers on node i. The first term inside the square brackets describes instances in which batches of ℓ negative customers are generated, but do not empty the node. The second term accounts for all transitions in which node i is emptied due to the generation of at least ℓ negative customers, (remembering that the excess negative customers have no effect on node i once it has been emptied). The LHS of equation (2.23) consists of all positive customer arrivals to node i, either from outside of the system, or as that proportion of the throughput of node j which is transferred to node i.

Equation (2.23) was obtained directly by interpreting the parameter Y_i to be the throughput of the single server at node i. It is a logical traffic equation describing the relationship between the throughputs of the servers in the network. We now show that it is equivalent to equation (2.9) which was obtained as the condition under which the network has a geometric invariant measure.

We can simplify each of the terms in square brackets on the RHS of equation (2.23). If we let $q_i = \left(\frac{Y_i}{\mu_i}\right)$ and consider the outside of the system to be node 0 then for $j \in \mathcal{L} \cup \{0\}$,

$$\begin{aligned}
&\sum_{\ell=1}^{\infty} \ell q_i^{\ell+1} d_{ji}^-(\ell) + \sum_{\ell=1}^{\infty} \ell(1-q_i) q_i^{\ell} \sum_{k=\ell}^{\infty} d_{ji}^-(k) \\
&\qquad = \sum_{\ell=1}^{\infty} \ell q_i^{\ell+1} d_{ji}^-(\ell) + q_i(1-q_i) \sum_{k=1}^{\infty} d_{ji}^-(k) \sum_{\ell=1}^{k} \ell q_i^{\ell-1}, \\
&\qquad = \sum_{\ell=1}^{\infty} \ell q_i^{\ell+1} d_{ji}^-(\ell) + q_i(1-q_i) \sum_{k=1}^{\infty} d_{ji}^-(k) \frac{d}{dq_i} \left[\sum_{\ell=1}^{k} q_i^{\ell} \right], \\
&\qquad = \sum_{\ell=1}^{\infty} \ell q_i^{\ell+1} d_{ji}^-(\ell) + q_i(1-q_i) \sum_{k=1}^{\infty} d_{ji}^-(k) \frac{d}{dq_i} \left[\frac{q_i - q_i^{k+1}}{1-q_i} \right], \\
&\qquad = \sum_{\ell=1}^{\infty} \ell q_i^{\ell+1} d_{ji}^-(\ell) + \frac{q_i}{1-q_i} \sum_{k=1}^{\infty} d_{ji}^-(k) \left[1 - (k+1) q_i^k + k q_i^{k+1} \right], \\
&\qquad = \frac{q_i}{1-q_i} \sum_{\ell=1}^{\infty} d_{ji}^-(\ell) \left[1 - q_i^l \right], \\
&\qquad = \sum_{\ell=1}^{\infty} d_{ji}^-(\ell) \sum_{k=1}^{\ell} q_i^k, \\
&\qquad = \sum_{\ell=1}^{\infty} q_i^{\ell} \sum_{k=\ell}^{\infty} d_{ji}^-(k).
\end{aligned}$$

Thus equation (2.23) becomes, for all $i \in \mathcal{L}$,

$$\sum_{j=0}^{N} Y_j d_{ji}^+ = Y_i + \sum_{j=0}^{N} Y_j \sum_{\ell=1}^{\infty} \left(\frac{Y_i}{\mu_i}\right)^{\ell} \sum_{k=\ell}^{\infty} d_{ji}^-(k),$$

which is equivalent to equation (2.9). That is, we have derived equation (2.9) from a logical traffic equation for throughputs.

Having interpreted Y_i as the throughput of the server at node i it is easy to see from equation (2.11) that, as stated by Gelenbe in [27] and [30], λ_i^+ is the arrival rate of positive customers to node i. Similarly, from equation (2.12), $\lambda_i^-(k)$ is the arrival rate of batches of k negative customers to node i. Finally, it is evident from equation (2.15) that the parameters g_i, which play a crucial role in the proof of Theorem 2.2.2, represent the probability that a positive customer arriving to node i will receive a complete service, and not be eliminated by a negative customer. This probability is the rate at which service completions occur at node i divided by the rate at which customers arrive to node i, namely Y_i divided by λ_i^+.

2.3 Partial Balance

With a geometric product form equilibrium distribution there is always a direct connection between traffic equations and partial balance equations, found by multiplying or dividing by $\pi(\mathbf{n})$. In this section we derive the partial balance equations for this network. These balance equations have natural interpretations, and are related to the effect that the boundaries of the state space have on the system.

Theorem 2.3.1 *Equation (2.3) satisfies the global balance equations (2.4) if and only if it satisfies the N partial balance equations, for $i \in \mathcal{L}$,*

$$\sum_{h=0}^{N} \pi(\mathbf{n}+\mathbf{e}_h-\mathbf{e}_i)\mu_h d_{hi}^{+} = \pi(\mathbf{n})\mu_i + \sum_{h=0}^{N}\sum_{\ell=1}^{\infty} \pi(\mathbf{n}+\mathbf{e}_h+(\ell-1)\mathbf{e}_i)\mu_h \sum_{k=\ell}^{\infty} d_{hi}^{-}(k) \quad (2.24)$$

and the additional partial balance equation

$$\begin{aligned}\pi(\mathbf{n})\sum_{i=1}^{N}\mu_0 d_{0i}^{+} \quad = \quad & \sum_{i=1}^{N}\Bigg[\pi(\mathbf{n}+\mathbf{e}_i)\mu_i d_{i0}^{+} + \sum_{h=1}^{N}\pi(\mathbf{n}+\mathbf{e}_h)\mu_h \sum_{\ell=1}^{\infty} d_{hi}^{-}(\ell) \\ & + \sum_{h=0}^{N}\sum_{\ell=1}^{\infty}\pi(\mathbf{n}+\mathbf{e}_h+\ell\mathbf{e}_i)\mu_h\sum_{k=\ell}^{\infty} d_{hi}^{-}(k)\Bigg]. \end{aligned} \quad (2.25)$$

Proof: We prove this using the same symbolic notation (A_i's and B_i's) as defined in Section 2.2.

$$\begin{aligned} \text{equation (2.9)} \quad &\Longleftrightarrow \quad \left(\tfrac{Y_i}{\mu_i}\right)(A_i+B_i)=0,\ \ i\in\mathcal{L}, \\ \text{equation (2.24)} \quad &\Longleftrightarrow \quad \pi(\mathbf{n})(A_i+B_i)=0, \\ \text{equation (2.25)} \quad &\Longleftrightarrow \quad \pi(\mathbf{n})\sum_{i=1}^{N} B_i = 0. \end{aligned} \quad (2.26)$$

From Theorem 2.2.1 we know that equation (2.3) satisfies the global balance equations (2.4) if and only if equation (2.9) holds. If the invariant measure of the network is given by equation (2.3), then equation (2.9) implies equation (2.24). Therefore, to complete the proof, we need to be able to derive the global balance equations (2.4), for $\mathcal{E}(\mathbf{n}) \subseteq \mathcal{L}$, from equations (2.24) and (2.25). This will ensure that equations (2.25) must also hold for the invariant measure to be of the form given by equation (2.3). Then, using the notation of (2.26), for $\mathcal{E}(\mathbf{n}) \subseteq \mathcal{L}$

$$
\begin{aligned}
& \sum_{i \in \mathcal{E}(\mathbf{n})^c} \Big(\text{equation } (2.24)\Big) \; - \; \text{equation } (2.25) \; = 0, \\
\iff & \quad \sum_{i \in \mathcal{E}(\mathbf{n})^c} \pi(\mathbf{n})(A_i + B_i) - \pi(\mathbf{n}) \sum_{i=1}^{N} B_i = 0, \\
\iff & \quad \pi(\mathbf{n})\Big(\sum_{i \in \mathcal{E}(\mathbf{n})^c} A_i - \sum_{i \in \mathcal{E}(\mathbf{n})} B_i \Big) = 0, \\
\iff & \quad \text{equation } (2.4).
\end{aligned}
$$

□

Referring to equations (2.24) and (2.25) as partial balance equations is not strictly accurate. Partial balance equations usually correspond to a simple separation of the terms of the global balance equations and are equivalent to considering global balance equations on an isolated subset of the sample space. This is not the case for the models of Gelenbe [27] and Gelenbe and Schassberger [31], nor for the model presented herein. The balance observed in each of these works is closer to, but not identical to, the equations derived by considering balance in and out of subsets of the state space.

Most partial balance concepts are clearer when explained in terms of probability flux which is relatively easy when nodes only have positive customers (see for example Whittle [74]). In our case it is a little more difficult but it does give some feel for the internal balance in this network and may provide sufficient insight to yield further generalisations.

Equations (2.24) and (2.25) are not independent. Equation (2.25) can be derived from equations (2.24) by replacing $\mathbf{n} - \mathbf{e}_i$ with $\mathbf{n}$ and then summing over i. Consequently an interpretation of these partial balance equations as flux equations is necessary only for equations (2.24).

The left hand side of equation (2.24) is the flux into state $\mathbf{n}$ due to the transfer of a positive customer to node i. The terms on the right hand side of equation (2.24) are respectively the flux out of state $\mathbf{n}$ due to a departure from node i and the flux due to the arrival of any type of customer to node i which leaves all other nodes in state n_j, $j \neq i$, and at the same time decreases node i from at least n_i to less than n_i customers. This flux balance is not as simple to explain in a few words as the flux balance for Jackson networks. The best approach that we can offer is via the concept of "source i interim states" which are $(N-1) \times 1$ vectors of the queue lengths observed, at nodes other than node i, *by any customer in transit to or from node i.* To this end define, for each $\mathbf{n} \in Z_+^N$, $\mathbf{n}(i) \in Z_+^{N-1}$ to be $\mathbf{n}(i) = (n_1, n_2, \ldots, n_{i-1}, n_{i+1}, \ldots, n_N)$.

For an arbitrary state $\mathbf{n}' \in Z_+^N$, and for a fixed $i \in \mathcal{L}$, equation (2.24) can be expressed as

> the probability flux balances between the sets of states $\{\mathbf{n} : n_i < n'_i\}$ and $\{\mathbf{n} : n_i \geq n'_i\}$ whenever the source i interim states have the property $\mathbf{n}(i) = \mathbf{n}'(i)$.

When there are no negative customers in the network this flux balance becomes the standard Whittle [74] balance for Jackson networks relating flux into state $\mathbf{n}$ due to an arrival to node i and the flux out of $\mathbf{n}$ due to a departure from node i.

2.4 Gelenbe's Results

In this section we show that the geometric product form equilibrium distribution derived by Gelenbe [30] for a network of queues with batch removal can be derived from Theorem 2.2.1. The networks in Gelenbe [27] are a special case of those in [30].

Gelenbe [30] considered an open network of N single server queues with mutually independent exponential service time distributions with rates $r(1), r(2), \ldots, r(N)$. Positive customers and batches of negative customers arrive to queue i according to independent Poisson processes with rates $\Lambda(i)$ and $\lambda(i)$ respectively. After completing its service at queue i, a positive customer will transfer to queue j as a positive customer with probability $p^+(i,j)$, as a batch of negative customers with probability $p^-(i,j)$, or it will depart the system with probability $d(i)$. As in our system, a customer is not allowed to transfer to the queue at which it was served. The size of a batch of negative customers is determined upon arrival to a node, and does not depend on the node of origin. A batch arriving to node i will consist of k negative customers with probability π_{ik}. Gelenbe showed that the geometric product form equilibrium distribution for this system exists with the following form

$$\pi(\mathbf{n}) = \prod_{i=1}^{N} (1 - q_i) q_i^{n_i}, \tag{2.27}$$

where

$$q_i = \frac{\lambda^+(i)}{r(i) + \lambda^-(i) f_i(q_i)}, \tag{2.28}$$

and

$$f_i(x) = \frac{1 - \sum_{k=1}^{\infty} \pi_{ik} x^k}{1 - x},$$

if and only if there exists unique non-negative $\lambda^+(i)$ and $\lambda^-(i)$ for $i \in \mathcal{L}$ that satisfy the following system of non-linear simultaneous equations

$$\lambda^+(i) = \sum_{j=1}^{N} q_j r(j) p^+(j,i) + \Lambda(i), \quad \lambda^-(i) = \sum_{j=1}^{N} q_j r(j) p^-(j,i) + \lambda(i). \tag{2.29}$$

This network can be obtained from our network by appropriately setting a few parameters. From our notation we require

$$\mu_0 d_{0i}^+ = \Lambda(i), \;\; \mu_0 d_{0i}^- = \lambda(i), \;\; \mu_i = r(i),$$

$$d_{ij}^+ = p^+(i,j), \;\; d_{ij}^-(k) = p^-(i,j)\pi_{jk}, \;\; d_{i0}^+ = d(i), \;\; \left(\frac{Y_i}{\mu_i}\right) = q_i.$$

Using these parameters equation (2.9) becomes

$$\begin{aligned}
q_i r(i) &= \sum_{j=1}^{N} q_j r(j) p^+(j,i) + \Lambda(i) - \sum_{j=1}^{N} q_j r(j) p^-(j,i) \sum_{\ell=1}^{\infty} q_i^\ell \sum_{k=\ell}^{\infty} \pi_{ik} - \lambda(i) \sum_{\ell=1}^{\infty} q_i^\ell \sum_{k=\ell}^{\infty} \pi_{ik}, \\
&= \lambda^+(i) - \sum_{\ell=1}^{\infty} q_i^\ell \sum_{k=\ell}^{\infty} \pi_{ik} \left(\sum_{j=1}^{N} q_j r(j) p^-(j,i) + \lambda(i) \right), \\
&= \lambda^+(i) - \lambda^-(i) \sum_{\ell=1}^{\infty} q_i^\ell \sum_{k=\ell}^{\infty} \pi_{ik}, \\
&= \lambda^+(i) - \lambda^-(i) \sum_{k=1}^{\infty} \pi_{ik} \sum_{\ell=1}^{k} q_i^\ell,
\end{aligned}$$

$$
\begin{aligned}
&= \lambda^+(i) - \lambda^-(i) \sum_{k=1}^{\infty} \pi_{ik} \left(\sum_{\ell=1}^{\infty} q_i^\ell - \sum_{\ell=k+1}^{\infty} q_i^\ell \right), \\
&= \lambda^+(i) - \lambda^-(i) \sum_{k=1}^{\infty} \pi_{ik} \left(\frac{q_i}{1-q_i} - \frac{q_i^{k+1}}{1-q_i} \right), \\
&= \lambda^+(i) - \lambda^-(i) \sum_{k=1}^{\infty} \pi_{ik} q_i \frac{1-q_i^k}{1-q_i}, \\
&= \lambda^+(i) - \lambda^-(i) q_i \frac{1-\sum_{k=1}^{\infty} \pi_{ik} q_i^k}{1-q_i}, \\
&= \lambda^+(i) - \lambda^-(i) q_i f_i(q_i),
\end{aligned}
$$

so that equation (2.28) is satisfied, and the results of Gelenbe [30] follow.

In this chapter we have extended the set of queueing networks that include negative customers and have a geometric product form equilibrium distribution. We showed that the global balance equations for these networks can be built up from a number of partial balance equations relating to the effect of the boundaries of the state space on the network, and that the partial balance equations are equivalent to a set of logical traffic equations.

Chapter 3

State Dependent Signalling

Since the introduction of product form queueing networks to the literature a number of authors have investigated networks with state dependent arrival and service intensities. Jackson networks [47] are an example of this in which the arrival and service intensities at a given node can depend on the number of customers at that node.

In a study of Markov population processes, Kingman [54] introduces transition rates which are a function of the current state of the process, rather than simply a function of the number of customers in a single queue. Population processes only allow single customer movement, so that the transitions of the process are given by

$$\begin{aligned} q(\mathbf{n}, \mathbf{n} + \mathbf{e}_i) &= \alpha_i(\mathbf{n}), \\ q(\mathbf{n}, \mathbf{n} - \mathbf{e}_i) &= \beta_i(\mathbf{n}), \\ q(\mathbf{n}, \mathbf{n} - \mathbf{e}_i + \mathbf{e}_j) &= \gamma_{ij}(\mathbf{n}). \end{aligned}$$

Kingman substitutes these expressions into the GBEs in order to find particular forms for the parameters α_i, β_i and γ_{ij} which ensure the process has a product form equilibrium distribution. He concludes that sufficient conditions for product form are for the process to be either reversible or satisfy the partial balance requirements of Whittle [74]. A reversible process is one in which $\pi(\mathbf{n})q(\mathbf{n}, \mathbf{n}') = \pi(\mathbf{n}')q(\mathbf{n}', \mathbf{n})$, $\forall \mathbf{n}, \mathbf{n}' \in S$, so that it satisfies detailed balance for all pairs of states $\mathbf{n}, \mathbf{n}' \in S$. The degree of achievable state dependence varies with the restrictions which are placed on the network.

Kingman notes that, for a closed reversible process, setting $\gamma_{ij}(\mathbf{n}) = \mu_i(n_i)p_{ij}\nu_j(n_j)$ is sufficient to obtain a product form equilibrium distribution, so that transfer rates may depend on the queue lengths at both the origin and destination nodes, but not on the queue lengths of any other nodes.

Kelly [50] extends Kingman's closed reversible process example to an open migration process with transition rates of the form

$$\begin{aligned} q(\mathbf{n}, \mathbf{n} + \mathbf{e}_i) &= \lambda_i \frac{\psi(\mathbf{n} + \mathbf{e}_i)}{\psi(\mathbf{n})}, \\ q(\mathbf{n}, \mathbf{n} - \mathbf{e}_i) &= \mu_i \frac{\phi(\mathbf{n})}{\phi(\mathbf{n} - \mathbf{e}_i)}, \\ q(\mathbf{n}, \mathbf{n} - \mathbf{e}_i + \mathbf{e}_j) &= p_{ij} \frac{\phi(\mathbf{n})}{\phi(\mathbf{n} - \mathbf{e}_i)} \frac{\psi(\mathbf{n} + \mathbf{e}_j)}{\psi(\mathbf{n})}. \end{aligned}$$

He assumes that there exists a solution $\{x_i,\ i \in \mathcal{L}\}$ to

$$x_i \left(\sum_{j=1}^{N} p_{ij} + \mu_i \right) = \sum_{j=1}^{N} x_j p_{ji} + \lambda_i, \quad i \in \mathcal{L},$$

and shows that the process will have a product form invariant measure of the form

$$\pi(\mathbf{n}) = \frac{\psi(\mathbf{n})}{\phi(\mathbf{n})} \prod_{i=1}^{N} x_i^{n_i}$$

if at least one of the following is true:

1. $\psi(\mathbf{n}) = 1,\ \forall \mathbf{n} \in S$,

2. $x_i p_{ij} = x_j p_{ji}$ and $\lambda_i = x_i \mu_i,\ \forall i, j$.

Kingman [54] and Kelly [50] both find that for particular processes, with state dependent arrival and service intensities, the process will have a product form equilibrium distribution if it is reversible. Whittle [76], Chapter 4, Theorem 8.1, provides a characterisation of reversible migration processes with state dependent intensities. A special case of this characterisation, presented in the lecture notes of Dr. W. Henderson, extends the results of Kingman [54] by allowing, in a closed network, transitions with intensities of the form

$$q(\mathbf{n} + \mathbf{e}_i, \mathbf{n} + \mathbf{e}_j) = \mu_i(\mathbf{n} + \mathbf{e}_i) p_{ij} \nu_j(\mathbf{n} + \mathbf{e}_j), \tag{3.1}$$

where

$$\mu_i(\mathbf{n}+\mathbf{e}_i) = \mu_i \frac{\psi(\mathbf{n})}{\phi(\mathbf{n}+\mathbf{e}_i)} \text{ and } \nu_j(\mathbf{n}+\mathbf{e}_j) = \lambda_j \frac{\xi(\mathbf{n}+\mathbf{e}_j)}{\kappa(\mathbf{n})}.$$

The closed migration process with these transfer intensities also has a product form equilibrium distribution. Note that the functions $\xi(\cdot)$ and $\kappa(\cdot)$ may be used to model blocking phenomena.

State dependent transitions may also occur in non-reversible product form networks. The networks of Henderson and Taylor [36] incorporate state dependent emission rates of the form

$$\frac{\psi(\mathbf{n}-\mathbf{a})\xi(\mathbf{a})}{\phi(\mathbf{n})},$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_N)$, in describing a product form queueing network with batch arrivals and batch services. In such networks, batches of customers may be released simultaneously, from a number of nodes, with a_i being the size of the batch released from node i. We discuss batch movement queueing networks in more detail in Chapter 5. Serfozo [70] independently derives a single movement product form network in which the emission rate of customers from node i, when the process is in state $\mathbf{n}$, is given by

$$\frac{\psi(\mathbf{n}-\mathbf{e}_i)}{\phi(\mathbf{n})}.$$

Henderson [41] introduces state dependent arrival and service intensities, similar to those of equation (3.1) and Serfozo [70], to the queueing networks with negative customers of Gelenbe [30] and Henderson, Northcote and Taylor [40]. In the networks of [41] customers are emitted from node $i \in \{0, 1, \ldots, N\}$ when the network is in state $\mathbf{n}$ at rate

$$\mu_i \frac{\psi(\mathbf{n}-\mathbf{e}_i)}{\phi(\mathbf{n})}.$$

Node 0 represents the outside of the network, so that the case $i = 0$ represents the rate at which customers arrive to the network. Customers are routed through the network as positive customers, or batches of k negative customers, the probability distributions of which are identical to those in Chapter 2, namely d_{ij}^{+} and $d_{ij}^{-}(k)$ respectively.

This chapter establishes a product form equilibrium distribution for networks of queues with state dependent arrival and service intensities, as in Henderson [41], and "signals". The term "signals" is used to describe a number of entities which affect/control network behaviour. The triggers of Gelenbe [29] and the interruptions of Henderson and Taylor [33] are referred to as type 0 signals and the batches of negative customers of [30], [40] and [41] are referred to as type t ($t \geq 1$) signals, where t is the batch size. We allow signals of these types to be generated within the network at the completion of a service time, and to arrive from outside the network.

The network under consideration is defined in Section 3.1 and analysed in Section 3.2 where it is shown, in particular, that the network has a product form equilibrium distribution. The special cases presented in Section 3.3, which describe networks with state space truncation and multiple customer classes for both signals and customers, are direct extensions of the work in [41], but are also applicable to the networks in this chapter, and are restated here in order to introduce the applications in Sections 3.4 to 3.7.

3.1 System Description

The network consists of N nodes, indexed as in Chapter 2 by a set of node labels $\mathcal{L} = \{1, 2, \ldots, N\}$. Associated with each node in the network is a queue of customers and a server/source. The term source is used because, unlike in Chapter 2, we allow customers and signals to be emitted from a node in the network even when the queue length at that node is zero or negative. In such situations the node is behaving as an infinite source rather than as a queue with a server. A customer emitted from an empty node will cause the queue length at the node to become negative. The state space, $S = Z^N$, of the network comprises vectors $\mathbf{n} = (n_1, n_2, \ldots, n_N)$ where $n_i \in Z$ is the queue length of customers at node i. We also consider there to be an external source of customers and signals, which we label node 0, representing the outside of the network. There is no queue of customers associated with node 0. Customers enter the network by being emitted from node 0 and leave by being absorbed by node 0.

Let $\mathbf{e}_i$ be an $N \times 1$ vector of zeros with a 1 in the ith position for $i \in \mathcal{L}$, and let $\mathbf{e}_0$ be the zero vector. Let $\phi(\cdot)$ be a real valued positive function on Z^N and $\psi(\cdot)$ be a real valued non-negative function defined on Z^N satisfying

$$\psi(\mathbf{n} - \ell \mathbf{e}_i) \geq \psi(\mathbf{n} - (\ell + 1)\mathbf{e}_i) \ \forall \ \mathbf{n}, \ i \in \mathcal{L}, \ \text{and} \ \ell \geq 0. \tag{3.2}$$

Note that $\mathbf{n} - \ell \mathbf{e}_i$ may include negative components.

3.1.1 Customer and Signal Emission

When the network is in state $\mathbf{n}$ source i, $i \in \mathcal{L} \cup \{0\}$ emits positive customers at rate

$$\mu_i \frac{\psi(\mathbf{n} - \mathbf{e}_i)}{\phi(\mathbf{n})}$$

causing the length of the queue at node i to decrease by one. For some applications negative queue lengths are acceptable. The source at node i may emit a customer even though there are no customers in the associated queue. In such circumstances the length of the queue at node i becomes negative. Note that node 0 will emit customers at rate

$$\mu_0 \frac{\psi(\mathbf{n} - \mathbf{e}_0)}{\phi(\mathbf{n})} = \mu_0 \frac{\psi(\mathbf{n})}{\phi(\mathbf{n})}.$$

Each emitted customer will undertake one of a number of courses of action according to a fixed probability distribution. A customer emitted from node i, for $i \in \mathcal{L} \cup \{0\}$, will

- be transferred to node $j \in \mathcal{L} \cup \{0\}$, causing the length of the queue at node j to increase by 1, (except at node 0), with probability d^+_{ij},
- become a type $t \geq 0$ signal and be transferred to node $j \in \mathcal{L}$ with probability $d^s_{ij}(t)$.

Thus, upon emission from source i a positive customer may leave the network, transfer to another node, or become a type t signal which moves to another node. To avoid duplication of intensities we assume that $d^+_{ii} = d^s_{i0}(t) = 0,\ i \in \mathcal{L} \cup \{0\},\ t \geq 0$, and so we require

$$d^+_{i0} + \sum_{j=1}^{N} \left[d^+_{ij} + \sum_{t=0}^{\infty} d^s_{ij}(t) \right] = 1, \text{ for } i \in \mathcal{L} \cup \{0\}.$$

Example: *The state dependence of the arrival and service intensities is incorporated in the functions* $\psi(\cdot) \geq 0$ *and* $\phi(\cdot) > 0$. *We may set these to be any real valued functions, providing we satisfy requirement (3.2). For example, to model a network of* $M|M|\infty$ *nodes (infinite server queues with Poisson arrivals), we set* $d^s_{ij}(t) = 0,\ \forall t$, *and*

$$\psi(\mathbf{n}) = \phi(\mathbf{n}) = \begin{cases} 0, & \text{if } n_i < 0 \text{ for any } i \in \mathcal{L}, \\ 1, & \text{if } \mathbf{n} = \mathbf{0}, \\ \prod_{i=1}^{N} \prod_{\ell=1}^{n_i} \ell^{-1}, & \text{otherwise,} \end{cases}$$

so that the rate of customer emission from node i *is*

$$\mu_i \frac{\psi(\mathbf{n} - \mathbf{e}_i)}{\phi(\mathbf{n})} = \mu_i n_i,$$

and the rate of customer arrivals to node $i,\ i \in \mathcal{L}$ *is* $\mu_0 d^+_{0i}$. *Additional examples of possible choices for* $\psi(\cdot)$ *and* $\phi(\cdot)$ *will be given in Sections 3.3 to 3.7.*

3.1.2 Signal Conversion at the Destination Node

Signals arriving to their destination node will be deemed either effective or ineffective by that node, and may also be converted into a different type of signal by that node. Each of these decisions is made according to a given state dependent probability distribution.

Upon arrival to node j, when the network is in state $\mathbf{n}$, signals of any type $t \geq 0$ are first of all deemed to be effective or ineffective with probabilities

$$\frac{\psi(\mathbf{n} - \mathbf{e}_j)}{\psi(\mathbf{n})} \text{ and } 1 - \frac{\psi(\mathbf{n} - \mathbf{e}_j)}{\psi(\mathbf{n})} \text{ respectively.}$$

An ineffective signal leaves the network in state $\mathbf{n}$. Thus, having been deemed to be ineffective, a signal has no further effect on the network. In many applications $\frac{\psi(\mathbf{n}-\mathbf{e}_j)}{\psi(\mathbf{n})}$ will be set to zero whenever the queue at node j is empty so that a signal arriving to an empty queue is always classed as an ineffective signal.

Effective type 0 signals play a similar role in the network to the interruptions of Henderson and Taylor [33] or the triggers of Gelenbe [29] in that they trigger the transfer of a customer from node j to node k with probability q_{jk}. We assume without loss of generality that a type 0 signal cannot trigger the transfer of a customer to the outside of the network, (these transitions are duplicated by type 1 signals), and so

$$\sum_{k=1}^{N} q_{jk} = 1. \tag{3.3}$$

Type t, $t \geq 1$ signals are analogous to a batch of t negative customers and arrive at node j (the destination node) with the *intention* of removing t customers from node j. An effective type t, $t \geq 1$, signal, arriving at node j with the network in interim state $\mathbf{n}$ decreases the queue length of queue j by ℓ (becomes a type ℓ signal) with probability

$$\frac{\psi(\mathbf{n}-t\mathbf{e}_j)}{\psi(\mathbf{n}-\mathbf{e}_j)}, \qquad \ell = t, \tag{3.4}$$

$$\frac{\psi(\mathbf{n}-\ell\mathbf{e}_j)-\psi(\mathbf{n}-(\ell+1)\mathbf{e}_j)}{\psi(\mathbf{n}-\mathbf{e}_j)}, \quad 1 \leq \ell < t. \tag{3.5}$$

The above probabilities are multiplied by $\frac{\psi(\mathbf{n}-\mathbf{e}_j)}{\psi(\mathbf{n})}$ to obtain the probabilities that a type t, $t \geq 1$ signal arriving at node j in state $\mathbf{n}$ is effective *and* removes ℓ customers from queue j. Equation (3.2) ensures that expressions (3.4) and (3.5) are non-negative. Note that

$$\frac{\psi(\mathbf{n}-t\mathbf{e}_j)}{\psi(\mathbf{n}-\mathbf{e}_j)} + \sum_{\ell=1}^{t-1} \frac{\psi(\mathbf{n}-\ell\mathbf{e}_j)-\psi(\mathbf{n}-(\ell+1)\mathbf{e}_j)}{\psi(\mathbf{n}-\mathbf{e}_j)} = 1,$$

ensuring that the probability distribution given by (3.4) and (3.5) is well defined, and that the probability that any signal of type $t \geq \ell \geq 1$ removes at least ℓ customers from queue j, when the system is in state $\mathbf{n}$, is

$$\frac{\psi(\mathbf{n}-\ell\mathbf{e}_j)}{\psi(\mathbf{n})}. \tag{3.6}$$

In summary, any signal arriving to a node has a state dependent probability of having an effect at that node. Type 0 signals cannot change type. Effective type 0 signals will always trigger the transfer of a customer from its destination node to another node (but not to node 0). An effective type t signal may be converted to a type ℓ ($1 \leq \ell \leq t$) signal by its destination node, triggering the immediate destruction of ℓ customers at the queue associated with that node. Ineffective signals correspond to a customer being served and leaving the network and therefore have no signalling effect on the network.

Henderson [41] uses similar signal conversion probabilities, which are essential for the network to have a product form invariant measure, due to the introduction of state dependent emission rates.

Remark 3.1.1 *Networks presented in other papers which include negative customers and triggers are special cases of the network that we have described. In particular we can describe the other networks using our notation in the following manner:*

- *The networks of [41] are duplicated by setting* $d^s_{ij}(0) = 0,\ \forall i, j$, *as they do not include type* 0 *signals.*
- *The networks of [30] and [40] can be found by setting* $d^s_{ij}(0) = 0,\ \forall i, j,\ \phi(\cdot) = 1$, $\psi(\mathbf{n}) = 1$ *if* $\mathbf{n} \in Z^N_+$ *and* $\psi(\mathbf{n}) = 0$ *otherwise, as they do not include type* 0 *signals, signal conversion or negative queue lengths.*
- *The networks of [29] are produced with* $d^s_{ij}(t) = 0$, $\forall i, j,\ t > 0,\ \phi(\cdot) = 1,\ \psi(\mathbf{n}) = 1$ *if* $\mathbf{n} \in Z^N_+$ *and* $\psi(\mathbf{n}) = 0$ *otherwise, as they do not include type* t, $t > 0$, *signals, or negative queue lengths.*

3.2 The Equilibrium Distribution

The global balance equations for the network can be constructed using the following rates into and out of state $\mathbf{n}$. Transitions out of state $\mathbf{n}$ occur when

- a customer is emitted from source $i \in \mathcal{L}$ at rate
$$\mu_i \frac{\psi(\mathbf{n} - \mathbf{e}_i)}{\phi(\mathbf{n})},$$

- a customer emitted from source 0 is transferred as an effective signal to node $i \in \mathcal{L}$ at rate
$$\mu_0 \frac{\psi(\mathbf{n})}{\phi(\mathbf{n})} \sum_{t=0}^{\infty} d_{0i}^s(t) \frac{\psi(\mathbf{n} - \mathbf{e}_i)}{\psi(\mathbf{n})},$$
- a customer emitted from source 0 is transferred to node $i \in \mathcal{L}$ at rate
$$\mu_0 \frac{\psi(\mathbf{n})}{\phi(\mathbf{n})} d_{0i}^+.$$

Transitions into state $\mathbf{n}$ occur from states

- $\mathbf{n} - \mathbf{e}_i$ when a customer is emitted from source 0 and transferred to node $i \in \mathcal{L}$ at rate
$$\mu_0 \frac{\psi(\mathbf{n} - \mathbf{e}_i)}{\phi(\mathbf{n} - \mathbf{e}_i)} d_{0i}^+,$$
- $\mathbf{n} + \mathbf{e}_i$ when a customer is emitted from source $i \in \mathcal{L}$ and transferred to node 0, or when an ineffective signal is emitted from source $i \in \mathcal{L}$ at rate
$$\mu_i \frac{\psi(\mathbf{n})}{\phi(\mathbf{n} + \mathbf{e}_i)} \left\{ d_{i0}^+ + \sum_{j=1}^{N} \sum_{t=0}^{\infty} d_{ij}^s(t) \frac{\psi(\mathbf{n}) - \psi(\mathbf{n} - \mathbf{e}_j)}{\psi(\mathbf{n})} \right\},$$
- $\mathbf{n} + \mathbf{e}_i - \mathbf{e}_k$ when a customer is emitted from source $i \in \mathcal{L}$ and transferred to node $k \in \mathcal{L}$ at rate
$$\mu_i \frac{\psi(\mathbf{n} - \mathbf{e}_k)}{\phi(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_k)} d_{ik}^+,$$
- $\mathbf{n} + \mathbf{e}_i + \mathbf{e}_j - \mathbf{e}_k$ when a type 0 signal is emitted from source $j \in \mathcal{L} \cup \{0\}$ and transferred to node $i \in \mathcal{L}$, triggering a customer transfer from node i to node $k \in \mathcal{L}$ at rate
$$\mu_j \frac{\psi(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_k)}{\phi(\mathbf{n} + \mathbf{e}_i + \mathbf{e}_j - \mathbf{e}_k)} d_{ji}^s(0) \frac{\psi(\mathbf{n} - \mathbf{e}_k)}{\psi(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_k)} q_{ik},$$
- $\mathbf{n} + \ell\mathbf{e}_i + \mathbf{e}_j$ when a type t ($t \geq \ell \geq 1$) signal is emitted from source $j \in \mathcal{L} \cup \{0\}$ and transferred to node $i \in \mathcal{L}$, triggering the destruction of ℓ customers at node i at rate
$$\mu_j \frac{\psi(\mathbf{n} + \ell\mathbf{e}_i)}{\phi(\mathbf{n} + \ell\mathbf{e}_i + \mathbf{e}_j)} \left\{ \sum_{t=\ell+1}^{\infty} d_{ji}^s(t) \frac{\psi(\mathbf{n}) - \psi(\mathbf{n} - \mathbf{e}_i)}{\psi(\mathbf{n} + \ell\mathbf{e}_i)} + d_{ji}^s(\ell) \frac{\psi(\mathbf{n})}{\psi(\mathbf{n} + \ell\mathbf{e}_i)} \right\}.$$

The state space S of the network is defined by the choice of the functions $\psi(\cdot)$ and $\phi(\cdot)$. This is demonstrated in Section 3.3, and further examples can be found in [41].

Theorem 3.2.1 *The network defined in Section 3.1 has the invariant measure*

$$\pi(\mathbf{n}) = K\phi(\mathbf{n})\prod_{i=1}^{N} x_i^{n_i}, \tag{3.7}$$

where $\{x_i > 0,\ i \in \mathcal{L}\}$ *satisfies, for* $i \in \mathcal{L}$,

$$\begin{aligned} \mu_i x_i \;=\; & \mu_0 d_{0i}^{+} + \sum_{h=1}^{N} \mu_h x_h \left[d_{hi}^{+} + \mu_0 d_{0h}^{s}(0) q_{hi} \frac{1}{\mu_h} + \sum_{k=1}^{N} x_k d_{hk}^{s}(0) q_{ki} \right] \\ & - \sum_{\ell=1}^{\infty} x_i^{\ell} \left[\mu_0 \sum_{t=\ell}^{\infty} d_{0i}^{s}(t) + \sum_{j=1}^{N} \mu_j x_j \sum_{t=\ell}^{\infty} d_{ji}^{s}(t) \right] \\ & - x_i \left[\mu_0 d_{0i}^{s}(0) + \sum_{j=1}^{N} \mu_j x_j d_{ji}^{s}(0) \right]. \end{aligned} \tag{3.8}$$

If the Markov process corresponding to the network is regular and if

$$\sum_{\mathbf{n}\in S} \phi(\mathbf{n}) \prod_{i=1}^{N} x_i^{n_i} < \infty, \tag{3.9}$$

then equation (3.7) is the equilibrium distribution for the network with

$$K^{-1} = \sum_{\mathbf{n}\in S} \phi(\mathbf{n}) \prod_{i=1}^{N} x_i^{n_i}. \tag{3.10}$$

Proof: We prove that equation (3.7), subject to equation (3.8), satisfies the global balance equations for each state in the state space S. The following global balance equations for the network are derived directly by considering the rates, given above, at which each of the possible transitions into and out of state $\mathbf{n}$ occurs.

$$\begin{aligned} & \pi(\mathbf{n}) \sum_{i=1}^{N} \left[\mu_i \frac{\psi(\mathbf{n}-\mathbf{e}_i)}{\phi(\mathbf{n})} + \mu_0 \frac{\psi(\mathbf{n})}{\phi(\mathbf{n})} \left\{ d_{0i}^{+} + \sum_{t=0}^{\infty} d_{0i}^{s}(t) \frac{\psi(\mathbf{n}-\mathbf{e}_i)}{\psi(\mathbf{n})} \right\} \right] \\ & \quad = \sum_{i=1}^{N} \Bigg[\pi(\mathbf{n}-\mathbf{e}_i) \mu_0 \frac{\psi(\mathbf{n}-\mathbf{e}_i)}{\phi(\mathbf{n}-\mathbf{e}_i)} d_{0i}^{+} \\ & \qquad + \pi(\mathbf{n}+\mathbf{e}_i) \mu_i \frac{\psi(\mathbf{n})}{\phi(\mathbf{n}+\mathbf{e}_i)} \left\{ d_{i0}^{+} + \sum_{j=1}^{N} \sum_{t=0}^{\infty} d_{ij}^{s}(t) \frac{\psi(\mathbf{n}) - \psi(\mathbf{n}-\mathbf{e}_j)}{\psi(\mathbf{n})} \right\} \\ & \qquad + \sum_{k=1}^{N} \left\{ \pi(\mathbf{n}+\mathbf{e}_i-\mathbf{e}_k) \mu_i \frac{\psi(\mathbf{n}-\mathbf{e}_k)}{\phi(\mathbf{n}+\mathbf{e}_i-\mathbf{e}_k)} d_{ik}^{+} \right. \end{aligned}$$

$$+\sum_{j=0}^{N}\pi(\mathbf{n}+\mathbf{e}_i+\mathbf{e}_j-\mathbf{e}_k)\mu_j\frac{\psi(\mathbf{n}+\mathbf{e}_i-\mathbf{e}_k)}{\phi(\mathbf{n}+\mathbf{e}_i+\mathbf{e}_j-\mathbf{e}_k)}d_{ji}^s(0)\frac{\psi(\mathbf{n}-\mathbf{e}_k)}{\psi(\mathbf{n}+\mathbf{e}_i-\mathbf{e}_k)}q_{ik}\Bigg\}$$
$$+\sum_{j=0}^{N}\sum_{\ell=1}^{\infty}\pi(\mathbf{n}+\ell\mathbf{e}_i+\mathbf{e}_j)\mu_j\frac{\psi(\mathbf{n}+\ell\mathbf{e}_i)}{\phi(\mathbf{n}+\ell\mathbf{e}_i+\mathbf{e}_j)}$$
$$\times\left\{\sum_{t=\ell+1}^{\infty}d_{ji}^s(t)\frac{\psi(\mathbf{n})-\psi(\mathbf{n}-\mathbf{e}_i)}{\psi(\mathbf{n}+\ell\mathbf{e}_i)}+d_{ji}^s(\ell)\frac{\psi(\mathbf{n})}{\psi(\mathbf{n}+\ell\mathbf{e}_i)}\right\}\Bigg]. \qquad (3.11)$$

Substituting for $\pi(\mathbf{n})$ from equation (3.7) we obtain

$$\sum_{i=1}^{N}\left[\mu_i\psi(\mathbf{n}-\mathbf{e}_i)+\mu_0\psi(\mathbf{n})d_{0i}^{+}+\mu_0\psi(\mathbf{n}-\mathbf{e}_i)\sum_{t=0}^{\infty}d_{0i}^s(t)\right]$$
$$=\sum_{i=1}^{N}\Bigg[\frac{1}{x_i}\mu_0\psi(\mathbf{n}-\mathbf{e}_i)d_{0i}^{+}+\mu_ix_i\psi(\mathbf{n})d_{i0}^{+}+\mu_ix_i\sum_{j=1}^{N}\sum_{t=0}^{\infty}d_{ij}^s(t)\left\{\psi(\mathbf{n})-\psi(\mathbf{n}-\mathbf{e}_j)\right\}$$
$$+\sum_{k=1}^{N}\Bigg\{\mu_ix_i\frac{1}{x_k}\psi(\mathbf{n}-\mathbf{e}_k)d_{ik}^{+}+\mu_0x_i\frac{1}{x_k}d_{0i}^s(0)\psi(\mathbf{n}-\mathbf{e}_k)q_{ik}$$
$$+\sum_{j=1}^{N}\mu_jx_jx_i\frac{1}{x_k}d_{ji}^s(0)\psi(\mathbf{n}-\mathbf{e}_k)q_{ik}\Bigg\}$$
$$+\mu_0\sum_{\ell=1}^{\infty}x_i^{\ell}\left\{d_{0i}^s(\ell)\psi(\mathbf{n})+\sum_{t=\ell+1}^{\infty}d_{0i}^s(t)(\psi(\mathbf{n})-\psi(\mathbf{n}-\mathbf{e}_i))\right\}$$
$$+\sum_{j=1}^{N}\mu_jx_j\sum_{\ell=1}^{\infty}x_i^{\ell}\left\{d_{ji}^s(\ell)\psi(\mathbf{n})+\sum_{t=\ell+1}^{\infty}d_{ji}^s(t)(\psi(\mathbf{n})-\psi(\mathbf{n}-\mathbf{e}_i))\right\}\Bigg].$$

Relabelling subscripts in summations and rearranging to get all $\psi(\mathbf{n}-\mathbf{e}_i)$ terms on the left hand side and all $\psi(\mathbf{n})$ terms on the right hand side we obtain

$$\sum_{i=1}^{N}\psi(\mathbf{n}-\mathbf{e}_i)\Bigg[\mu_i+\mu_0\sum_{t=0}^{\infty}d_{0i}^s(t)-\frac{1}{x_i}\mu_0d_{0i}^{+}+\sum_{j=1}^{N}\mu_jx_j\sum_{t=0}^{\infty}d_{ji}^s(t)$$
$$-\sum_{k=1}^{N}\left\{\mu_kx_k\frac{1}{x_i}d_{ki}^{+}+\mu_0x_k\frac{1}{x_i}d_{0k}^s(0)q_{ki}+\sum_{j=1}^{N}\mu_jx_jx_k\frac{1}{x_i}d_{jk}^s(0)q_{ki}\right\}$$
$$+\mu_0\sum_{\ell=1}^{\infty}x_i^{\ell}\sum_{t=\ell+1}^{\infty}d_{0i}^s(t)+\sum_{j=1}^{N}\mu_jx_j\sum_{\ell=1}^{\infty}x_i^{\ell}\sum_{t=\ell+1}^{\infty}d_{ji}^s(t)\Bigg]$$
$$=\psi(\mathbf{n})\sum_{i=1}^{N}\Bigg[\mu_ix_id_{i0}^{+}+\mu_ix_i\sum_{j=1}^{N}\sum_{t=0}^{\infty}d_{ij}^s(t)+\mu_0\sum_{\ell=1}^{\infty}x_i^{\ell}\sum_{t=\ell}^{\infty}d_{0i}^s(t)$$
$$+\sum_{j=1}^{N}\mu_jx_j\sum_{\ell=1}^{\infty}x_i^{\ell}\sum_{t=\ell}^{\infty}d_{ji}^s(t)-\mu_0d_{0i}^{+}\Bigg].$$

To complete the proof we show that

$$\text{coefficient of } \psi(\mathbf{n}-\mathbf{e}_i)=0, \qquad (3.12)$$

and

$$\text{coefficient of } \psi(\mathbf{n}) = 0. \tag{3.13}$$

Equations (3.12) and (3.13) are traffic equations, which can be shown to correspond to partial balance equations, and are sufficient for the network to have product form decomposability.

coefficient of $\psi(\mathbf{n}-\mathbf{e}_i) \times x_i$

$$\begin{aligned}
&= \mu_i x_i + x_i \mu_0 \sum_{t=0}^{\infty} d_{0i}^s(t) - \mu_0 d_{0i}^+ + x_i \sum_{j=1}^{N} \mu_j x_j \sum_{t=0}^{\infty} d_{ji}^s(t) - \sum_{k=1}^{N} \mu_k x_k d_{ki}^+ \\
&\quad - \sum_{k=1}^{N} \mu_0 x_k d_{0k}^s(0) q_{ki} - \sum_{k=1}^{N}\sum_{j=1}^{N} \mu_j x_j x_k d_{jk}^s(0) q_{ki} \\
&\quad + \mu_0 \sum_{\ell=1}^{\infty} x_i^{\ell+1} \sum_{t=\ell+1}^{\infty} d_{0i}^s(t) + \sum_{j=1}^{N} \mu_j x_j \sum_{\ell=1}^{\infty} x_i^{\ell+1} \sum_{t=\ell+1}^{\infty} d_{ji}^s(t), \\
&= \mu_i x_i - \mu_0 d_{0i}^+ - \sum_{h=1}^{N} \mu_h x_h \left[d_{hi}^+ + \mu_0 d_{0h}^s(0) q_{hi} \frac{1}{\mu_h} + \sum_{k=1}^{N} x_k d_{hk}^s(0) q_{ki} \right] \\
&\quad + \sum_{\ell=0}^{\infty} x_i^{\ell+1} \left[\mu_0 \sum_{t=\ell+1}^{\infty} d_{0i}^s(t) + \sum_{j=1}^{N} \mu_j x_j \sum_{t=\ell+1}^{\infty} d_{ji}^s(t) \right] + x_i \left[\mu_0 d_{0i}^s(0) + \sum_{j=1}^{N} \mu_j x_j d_{ji}^s(0) \right], \\
&= \mu_i x_i - \mu_0 d_{0i}^+ - \sum_{h=1}^{N} \mu_h x_h \left[d_{hi}^+ + \mu_0 d_{0h}^s(0) q_{hi} \frac{1}{\mu_h} + \sum_{k=1}^{N} x_k d_{hk}^s(0) q_{ki} \right] \\
&\quad + \sum_{\ell=1}^{\infty} x_i^{\ell} \left[\mu_0 \sum_{t=\ell}^{\infty} d_{0i}^s(t) + \sum_{j=1}^{N} \mu_j x_j \sum_{t=\ell}^{\infty} d_{ji}^s(t) \right] + x_i \left[\mu_0 d_{0i}^s(0) + \sum_{j=1}^{N} \mu_j x_j d_{ji}^s(0) \right], \\
&= 0 \text{ by equation (3.8)},
\end{aligned}$$

and

coefficient of $\psi(\mathbf{n})$

$$\begin{aligned}
&= \sum_{i=1}^{N} \left[\mu_i x_i d_{i0}^+ + \mu_i x_i \sum_{j=1}^{N}\sum_{t=0}^{\infty} d_{ij}^s(t) + \mu_0 \sum_{\ell=1}^{\infty} x_i^{\ell} \sum_{t=\ell}^{\infty} d_{0i}^s(t) \right. \\
&\quad \left. + \sum_{j=1}^{N} \mu_j x_j \sum_{\ell=1}^{\infty} x_i^{\ell} \sum_{t=\ell}^{\infty} d_{ji}^s(t) - \mu_0 d_{0i}^+ \right], \\
&= \sum_{i=1}^{N} \left[\mu_i x_i \left\{ 1 - \sum_{j=1}^{N} d_{ij}^+ \right\} - \mu_0 d_{0i}^+ + \sum_{\ell=1}^{\infty} x_i^{\ell} \left\{ \mu_0 \sum_{t=\ell}^{\infty} d_{0i}^s(t) + \sum_{j=1}^{N} \mu_j x_j \sum_{t=\ell}^{\infty} d_{ji}^s(t) \right\} \right], \\
&= \sum_{i=1}^{N} \left[\sum_{h=1}^{N} \mu_h x_h \left\{ \mu_0 d_{0h}^s(0) q_{hi} \frac{1}{\mu_h} + \sum_{k=1}^{N} x_k d_{hk}^s(0) q_{ki} \right\} \right.
\end{aligned}$$

$$-x_i \left\{ \mu_0 d^s_{0i}(0) + \sum_{j=1}^{N} \mu_j x_j d^s_{ji}(0) \right\} \Bigg] \text{ by equation (3.8),}$$
$$= \quad 0 \text{ by equation (3.3).}$$

The theory of Markov processes (see, for example, [64]) ensures that if there is a solution $\pi(\mathbf{n})$ to equations (3.11) then it is unique up to a constant multiple. Thus any solution $\mu_i x_i$ to equation (3.8) must be unique. If the process describing the evolution of the network is regular then it is known that any normalised solution of the GBEs is an equilibrium distribution (see, for example, Miller [55]). Thus if condition (3.9) holds, with $\{0 < x_i < 1, \ \forall i \in \mathcal{L}\}$, then (3.7) with K defined by (3.10) is the equilibrium distribution for the network.

□

Remark 3.2.1 *The functions $\psi(\cdot)$ and $\phi(\cdot)$ dictate the speeds of the various intensities within the network, which may be state dependent. It is clear that the equilibrium behaviour of the network should not be affected by the speed of the arrival and service intensities. For this reason, the functions $\psi(\cdot)$ and $\phi(\cdot)$ do not affect the stability or equilibrium behaviour of the network. Stability only requires the parameters x_i to satisfy $0 < x_i < 1, \ \forall i \in \mathcal{L}$.*

Remark 3.2.2 *In Chapter 2, equations (2.3) and (2.9), corresponding to equations (3.7) and (3.8), are written in terms of the variables $Y_i = \mu_i x_i$, which are shown to have interpretations as the throughput of the source at node i. Although not explicitly stated in Chapter 2 this restricts consideration to the case $\mu_i > 0$, since for stability it is required that $0 \leq Y_i < \mu_i$. There are many practical applications in which the source at node i never emits customers, and the customers queued at node i are released only by signals. Such situations are modelled by setting $\mu_i = 0$. Recasting the equilibrium distribution in the form (3.7) and the "traffic equations" in the form (3.8) allows this* *case to be considered. We still interpret each $Y_i = \mu_i x_i$ in this chapter to be the throughput of the single server at source i.*

Theorem 3.2.2 *There exists a unique non negative solution $\{x_i,\ i \in \mathcal{L}\}$ to equations (3.8).*

Proof: Gelenbe [29] proves the existence of a solution for the traffic equations of a network which includes only type 0 signals (triggers), and Theorem 2.2.2 confirms the existence of a solution for a network which includes only type $t \geq 1$ signals. These two proofs are combined in the following to prove the existence of a solution $\{x_i,\ i \in \mathcal{L}\}$ to equations (3.8). Using similar notation to the proof of Theorem 2.2.2, and setting $x_0 = 1$, equations 3.8 become

$$\mu_i x_i = \lambda_i^+ - x_i \lambda_i^t - \sum_{\ell=1}^{\infty} x_i^\ell \sum_{k=\ell}^{\infty} \lambda_i^-(k),\ i \in \mathcal{L}, \tag{3.14}$$

where

$$\lambda_i^t = \sum_{j=0}^{N} \mu_j x_j d_{ji}^s(0),\ i \in \mathcal{L}, \tag{3.15}$$

$$\lambda_i^-(k) = \sum_{j=0}^{N} \mu_j x_j d_{ji}^-(k),\ i \in \mathcal{L},\ k \geq 1, \tag{3.16}$$

and

$$\begin{aligned} \lambda_i^+ &= \sum_{j=0}^{N} \mu_j x_j \left\{ d_{ji}^+ + \sum_{k=1}^{N} x_k d_{jk}^s(0) q_{ki} \right\},\ i \in \mathcal{L}, \\ &= \sum_{j=1}^{N} x_j \left\{ \mu_j d_{ji}^+ + \lambda_j^t q_{ji} \right\} + \mu_0 d_{0i}^+,\ i \in \mathcal{L}, \\ &= \sum_{j=1}^{N} \lambda_j^+ \left\{ g_j d_{ji}^+ + h_j q_{ji} \right\} + \mu_0 d_{0i}^+,\ i \in \mathcal{L}, \end{aligned} \tag{3.17}$$

where $\lambda_j^+ g_j = \mu_j x_j$ and $\lambda_j^+ h_j = \lambda_j^t x_j$. From equation (3.14)

$$x_i = \frac{\lambda_i^+}{\mu_i + \lambda_i^t + \sum_{k=1}^{\infty} \lambda_i^-(k) f_k(x_i)},\ i \in \mathcal{L}, \tag{3.18}$$

where

$$f_k(x) = \sum_{\ell=1}^{k} x^{\ell-1},\ k \geq 1.$$

Rewrite equation (3.14) as a polynomial in x_i.

$$F_i(x_i) = \mu_i x_i - \lambda_i^+ + x_i \lambda_i^t + \sum_{\ell=1}^{\infty} x_i^\ell \sum_{k=\ell}^{\infty} \lambda_i^-(k), \ i \in \mathcal{L}. \tag{3.19}$$

For each $i \in \mathcal{L}$, given fixed non-negative values of λ_i^+, λ_i^t and $\lambda_i^-(k)$, $k \geq 1$, x_i may be determined uniquely from equation (3.19) as a root of the polynomial $F_i(\cdot)$. This may be shown by noting the following and applying the mean value theorem and intermediate value theorem,

$$F_i(0) \leq 0,$$
$$\lim_{t\to\infty} F_i(t) > 0,$$
$$F_i'(x) > 0, \ x \geq 0.$$

Thus, for each $i \in \mathcal{L}$, given fixed non-negative values of λ_i^+, λ_i^t and $\lambda_i^-(k)$, $k \geq 1$, x_i, g_i and h_i may be determined uniquely. Equations (3.15) and (3.16) become

$$\begin{aligned} \lambda_i^t &= \sum_{j=1}^{N} \lambda_j^+ g_j d_{ji}^s(0) + \mu_0 d_{0i}^s(0), \ i \in \mathcal{L}, \\ \lambda_i^-(k) &= \sum_{j=1}^{N} \lambda_j^+ g_j d_{ji}^-(k) + \mu_0 d_{0i}^s(k), \ i \in \mathcal{L}, \ k \geq 1, \end{aligned}$$

so that we have the vector equations

$$\boldsymbol{\lambda}^t = \boldsymbol{\lambda}^+ G D^t + \boldsymbol{\mu}^t, \tag{3.20}$$

$$\boldsymbol{\lambda}^-(k) = \boldsymbol{\lambda}^+ G D_k^- + \boldsymbol{\mu}_k^-, \ k \geq 1, \tag{3.21}$$

$$\boldsymbol{\lambda}^+ = \boldsymbol{\lambda}^+ (G D^+ + H Q) + \boldsymbol{\mu}^+, \tag{3.22}$$

where $\boldsymbol{\lambda}^t$, $\boldsymbol{\lambda}^-(k)$, $\boldsymbol{\lambda}^+$, $\boldsymbol{\mu}^t$, $\boldsymbol{\mu}_k^-$, and $\boldsymbol{\mu}^+$ are $1 \times N$ row vectors whose i-th entries are λ_i^t, $\lambda_i^-(k)$, λ_i^+, $\mu_0 d_{0i}^t$, $\mu_0 d_{0i}^-(k)$, and $\mu_0 d_{0i}^+$ respectively and G, D^t, D_k^-, D^+, H and Q are $N \times N$ matrices given by

$$D^t = \left[d_{ij}^s(0)\right]_{ij}, \ D_k^- = \left[d_{ij}^-(k)\right]_{ij}, \ D^+ = \left[d_{ij}^+\right]_{ij},$$
$$G = diag(g_i), \ H = diag(h_i), \text{ and } Q = [q_{ij}]_{ij}.$$

Equations (3.21) may be combined by forming

$$\boldsymbol{\lambda}^- = \left[\boldsymbol{\lambda}^-(1), \boldsymbol{\lambda}^-(2), \ldots\right], \ D^- = \left[D_1^- \ D_2^- \ \ldots\right] \text{ and } \boldsymbol{\mu}^- = \left[\boldsymbol{\mu}_1^-, \boldsymbol{\mu}_2^-, \ldots\right],$$

so that

$$\boldsymbol{\lambda}^- = \boldsymbol{\lambda}^+ G D^- + \boldsymbol{\mu}^-. \tag{3.23}$$

From equation (3.22)

$$\boldsymbol{\lambda}^+(I - GD^+ - HQ) = \boldsymbol{\mu}^+.$$

Noting that $G + H \leq I$ component-wise, from equation (3.14) with $g_i + h_i = \frac{\mu_i x_i + \lambda_i^t x_i}{\lambda_i^+}$, that D^+ is sub-stochastic and Q is stochastic, each without any ergodic classes, then $(I - GD^+ - HQ)$ must be sub-stochastic. Therefore we can solve for $\boldsymbol{\lambda}^t$, $\boldsymbol{\lambda}^-$ and $\boldsymbol{\lambda}^+$ explicitly to get

$$\boldsymbol{\lambda}^+ = \boldsymbol{\mu}^+ \sum_{\ell=0}^{\infty} (GD^+ + HQ)^\ell, \tag{3.24}$$

$$\boldsymbol{\lambda}^t = \boldsymbol{\mu}^+ \sum_{\ell=0}^{\infty} (GD^+ + HQ)^\ell GD^t + \boldsymbol{\mu}^t, \tag{3.25}$$

$$\boldsymbol{\lambda}^- = \boldsymbol{\mu}^+ \sum_{\ell=0}^{\infty} (GD^+ + HQ)^\ell GD^- + \boldsymbol{\mu}^-. \tag{3.26}$$

We combine equations (3.24) (3.25) and (3.26), by setting $\boldsymbol{\lambda} = \left[\boldsymbol{\lambda}^+, \boldsymbol{\lambda}^t, \boldsymbol{\lambda}^-\right]$ and $\boldsymbol{\mu} = [\mathbf{0}, \boldsymbol{\mu}^t, \boldsymbol{\mu}^-]$, to obtain

$$\boldsymbol{\lambda} = \boldsymbol{\mu}^+ \sum_{\ell=0}^{\infty} (GD^+ + HQ)^\ell \left[I \quad GD^t \quad GD^-\right] + \boldsymbol{\mu}. \tag{3.27}$$

Let $\mathbf{m} = \boldsymbol{\lambda} - \boldsymbol{\mu}$ and define the function $F(\cdot)$ to be

$$F(\mathbf{m}) = \boldsymbol{\mu}^+ \sum_{\ell=0}^{\infty} (GD^+ HQ)^\ell \left[I \quad GD^t \quad GD^-\right] + \boldsymbol{\mu}, \tag{3.28}$$

where $F(\cdot)$ depends on $\mathbf{x}$ through the matrices G and H. For non-negative vectors $\mathbf{m}$, $F(\mathbf{m})$ is a non-negative continuous function and is maximised when $G + H = I$, since all other terms in equation (3.28) are non-negative constants. Let this maximum value of $F(\cdot)$, corresponding to $G + H = I$, be F^*, so that $F(\cdot)$ satisfies

$$F : [\mathbf{0}, F^*] \longrightarrow [\mathbf{0}, F^*].$$

Thus, by Brouwer's fixed point theorem [22], there exists a fixed point $\mathbf{m}^*$ satisfying $\mathbf{m}^* = F(\mathbf{m}^*)$. From this fixed point we can obtain fixed vectors $\boldsymbol{\lambda}^t$, $\boldsymbol{\lambda}^-$, $\boldsymbol{\lambda}^+$ and ultimately $\mathbf{x} = (x_1, x_2, \ldots, x_N)$.

□

3.3 Special Cases

3.3.1 State Space Truncation

Boucherie and van Dijk [6] and Henderson [41] discuss how the state spaces for queueing networks with negative customers may be given upper and lower bounds. These results may be extended to queueing networks with signals. The standard state space boundaries for queueing networks are $n_i = 0,\ i \in \mathcal{L}$. Following the approach used in Henderson [41] these can be incorporated into our network through the choice of the $\psi(\cdot)$ function. Setting $\psi(\mathbf{n}) = 0$ if $\mathbf{n} \leq \mathbf{0}$ will ensure that $\psi(\mathbf{n} - \ell\mathbf{e}_i) = 0,\ \forall \ell > n_i$. Thus type t $(t > n_i)$ signals arriving to node i will be transformed into type ℓ signals for $1 \leq \ell \leq n_i$ since, by this choice of $\psi(\cdot)$, there is a zero probability that the signal will become a type ℓ signal for $\ell > n_i$. Thus an arriving signal can at most empty the queue, maintaining the integrity of the state space boundaries. Similarly, when $n_i = 0$, type 0 signals will always be made ineffective by node i, and source i will not be able to emit any customers or signals since $\psi(\mathbf{n} - \mathbf{e}_i) = 0$.

Other boundaries can be created by appropriate choices of $\psi(\cdot)$. For example setting $\psi(\mathbf{n}) = 0$ for some $\mathbf{n}$ deletes all states $\mathbf{n}' \leq \mathbf{n}$ from the state space. In this way lower boundaries to the state space can be created.

To enforce an upper boundary on the state space set

$$\psi(\mathbf{n}) = \psi'(\mathbf{n})\phi(\mathbf{n}),$$

$$\phi(\mathbf{n}) = \prod_{i=1}^{N} \prod_{r=1}^{n_i} \left[\xi_i(r)\right]^{-1}.$$

The emission rate from source i, in this case, is

$$\mu_i \frac{\psi(\mathbf{n} - \mathbf{e}_i)}{\phi(\mathbf{n})} = \mu_i \psi'(\mathbf{n} - \mathbf{e}_i)\xi_i(n_i).$$

We can choose $\xi_i(\cdot)$ to be very large for all states outside of the required state space. For example set $\xi_i(\mathbf{n}) = \infty,\ \forall \mathbf{n} : n_i = N_i + 1$ where N_i is the upper limit of the queue length of source i. This forces a customer to be served at node i, decreasing the queue length by one, whenever the queue length exceeds N_i. This released customer is then routed through the network until it creates a state which is acceptable. The customer is behaving as if it is being blocked, but rather than returning to its previous queue it searches along its route to find service. The function $\psi'(\cdot)$ can again be chosen in any way we wish and, in particular, can be used, as before, to create lower boundaries for the state space.

3.3.2 Network Partitions and Multiple Customer Classes

In this section we show, as in Henderson [41], that a particular choice of $\phi(\cdot)$ allows the network to be partitioned so that the emission rates depend on the partitioning used. In particular we show that multiple customer classes may be introduced to our network as a special case of this partitioning method.

Partition the set of node labels, $\mathcal{L}$, in K distinct fashions so that the k-th, $1 \leq k \leq K$, partitioning is defined by the L_k mutually exclusive and exhaustive subsets of $\mathcal{L}$, denoted $\{A_\ell(k),\, \ell = 1, 2, \ldots, L_k\}$. Thus, for $k = 1, 2, \ldots, K$,

$$A_\ell(k) \subseteq \mathcal{L},$$

such that

$$\bigcup_{\ell=1}^{L_k} A_\ell(k) = \mathcal{L},$$

and

$$A_\ell(k) \cap A_{\ell'}(k) = \emptyset, \quad \forall\, \ell \neq \ell'.$$

Now define

$$\psi(\mathbf{n}) = \psi'(\mathbf{n})\phi(\mathbf{n}), \quad \forall \mathbf{n} \in S,$$

with

$$\phi(\mathbf{n}) = \prod_{k=1}^{K} \prod_{\ell=1}^{L_k} \prod_{r=1}^{m(A_\ell(k))} \left[\xi_{\ell,k}(r)\right]^{-1},$$

where $m(A_\ell(k))$ is the sum of the queue lengths of all queues in the set $A_\ell(k)$ when the state of the network is $\mathbf{n}$.

The emission rate of customers at node i, in this case, is

$$\mu_i \frac{\psi(\mathbf{n}-\mathbf{e}_i)}{\phi(\mathbf{n})} = \mu_i \psi'(\mathbf{n}-\mathbf{e}_i) \prod_{k=1}^{K} \prod_{\ell : i \in A_\ell(k)} \xi_{\ell,k}\Big(m(A_\ell(k))\Big). \tag{3.29}$$

Notice that there is only one ℓ satisfying $i \in A_\ell(k)$ for a particular i and k, so that the second product in equation (3.29) consists of only one term. We write it as a product to identify the term as succinctly as possible.

The component $\xi_{\ell,k}\Big(m(A_\ell(k))\Big)$ of this emission rate is common to every node in $A_\ell(k)$ and is a function of the total queue size of the queues of $A_\ell(k)$. Consequently the emission rate is made up of components contributed from each partition of the nodes. If the kth partition involves N sets, each containing a single node, then this partition contributes a component $\xi_i(n_i)$, as in Section 3.3.1, to the emission rate. As in Section 3.3.1, $\xi_i(n_i)$ can be chosen so as to limit the queue size of queue i to an upper bound. Using exactly the same procedure with the partitions above we can set $\xi_{\ell,k}\Big(m(A_\ell(k))\Big)$ to be very large to limit the joint queue size in the set of queues $A_\ell(k)$ to a particular value and, by the appropriate choice of partitions, can do so concurrently for arbitrary combinations of queues. When the service rate gets very large it is a little more difficult to be precise about which customer is instantly served, and from which queue in $A_\ell(k)$, without introducing queue disciplines, but it can be done in some situations (see for example Kelly [52]).

The function $\psi'(\cdot)$ can again be chosen in any way we wish and, in particular, can be used, as in Section 3.3.1, to create lower boundaries for the state space.

This method of partitioning the queues in the network can be used to analyse queueing networks with multiple customer classes and signals. Consider a network of $N \times C$ sub-nodes in which each sub-node is labelled by both a class of customer from the set $\{1, 2, \ldots, C\}$ and a node from the set $\mathcal{L}$. Sub-node $c\ell$ consists only of customers of class c which are associated with node ℓ. By direct analogy with the network defined in Section 3.1 the routeing probabilities for the $N \times C$ sub-node network are $d^+(c\ell; 0)$, $d^+(c\ell; dk)$ and $d^s(c\ell; dk; t)$ corresponding to d^+_{i0}, d^+_{ij} and $d^s_{ij}(t)$

respectively, with

$$d^{+}(c\ell;0) + \sum_{d=1}^{C}\sum_{k=1}^{N}\left[d^{+}(c\ell;dk) + \sum_{t=0}^{\infty} d^{s}(c\ell;dk;t)\right] = 1, \text{ for } 1 \le c \le C,\ 1 \le \ell \le N.$$

Thus customers can change class as they change nodes. Now define a single partition $\{A_\ell,\ \ell \in \mathcal{L}\}$ so that A_ℓ contains all sub-nodes associated with node ℓ. The queue at node ℓ is the aggregation of the C sub-queues incorporated in A_ℓ. Thus the number of customers in the queue at node ℓ is $m(A_\ell)$. The service rate offered to class c customers at node ℓ (customers at sub-node $c\ell$) is, from equation (3.29),

$$\mu_{c\ell}\frac{\psi(\mathbf{n}-\mathbf{e}_{c\ell})}{\phi(\mathbf{n})} = \mu_{c\ell}\psi'(\mathbf{n}-\mathbf{e}_{c\ell})\xi_\ell\Big(m(A_\ell)\Big). \tag{3.30}$$

It is a function of the total number of customers of all classes at node ℓ as well as being dependent on c which is the class of customer represented at sub-node $c\ell$. Consequently customers change nodes and classes according to the above probability distribution. They request a mean service time that depends on both their class and current node. The service facility at each node works at a rate dependent upon the total number of customers at the node, as in Jackson and BCMP networks, but does not distinguish between customer classes. These are the standard assumptions made when different classes of customers are circulating in the network and therefore, through this example, we have analysed networks of queues with different classes of customers and different types of signals moving between nodes which serve customers with state dependent intensities.

Chandy and Martin [13] and Kelly [52] show that to have class dependent mean service times at a queue in a product form queueing network the queue had to be a symmetric queue. Equation (3.30) allows class dependent mean service times apparently without restricting the queue discipline in any way. At first glance this appears to be a contradiction. However in compacting the $C \times N$ sub-node network into an N node network an implicit assumption has been made that the service facility at each node does not favour any particular class of customer. This fact, which can be established rigorously, is equivalent to the assumption of processor sharing, (an example of a symmetric discipline), and thus no contradiction exists.

3.4 Manufacturing, Supply and Demand Models

Consider a network of two nodes, with multiple classes of both customers and signals, which we will use to model the manufacture, storage and distribution of a particular product. Our model may be used to determine congestions or deficiencies in these processes, and may also be used to determine optimal operating conditions for each stage in the process.

We consider the manufacture of the product to occur at a single node, (node s say), in X stages. For a particular unit of the product, stage $x+1$ in the manufacturing process may not begin until stage x has been completed. Let n_{sx} be the number of units which have passed through the first x stages of production and are queued at node s for stage $x+1$. The raw materials required for manufacture of a single unit arrive to node s together, with rate parameter μp_s, and are stored as a single class 0 customer. This implies that at any time there will be raw materials available to start production on n_{s0} units of the product. The class of a customer will increase by one each time it completes a production stage, so that a class i, $i < X$ customer will become a class $i+1$ customer once it has passed through stage i of the manufacturing process. This is modelled by setting $d^+(sx, s(x+1)) = 1$ for $x < X$. Class X customers have passed through all of the production stages, so that n_{sX} is the number of completed units waiting to be distributed. If $\frac{1}{\mu_s}$ is the mean time taken for a single unit to pass through the whole manufacturing process (without queueing anywhere) and ρ_x is the proportion of that time spent in the x-th stage of production then $\mu_{sx} = \frac{\mu_s}{\rho_x}$ is the rate parameter at which class x customers are processed to become class $x+1$ customers.

The storage and distribution processes for the product are only concerned with class X customers. We consider all demand for the product to aggregate at a single node, (node d say). Orders arrive to and are queued at node d, with rate parameter μp_d, as class X customers. These orders are processed with rate parameter μ_{dX}. An order will request t units of the completed product with probability $r(t)$. Each order can be thought of as a request for t units of the product, with t given by the probability distribution $r(t)$. Let $p_s + p_d = 1$, so μ is the aggregate rate parameter for all arrivals to the system. When the supplier, (node s), satisfies a request for a unit of the product, the level of stocks in the supplier's warehouse decreases by one unit (n_{sX} decreases by one). We add a little more complexity to this system by assuming that the supplier of the product may only satisfy a request if there are more than R (some constant) units of the product available. If there are n_{sX} ($R < n_{sX} \leq R + t$) units in the warehouse and an order for t units arrives to the supplier, then only the first $n_{sX} - R$ unit requests will be satisfied, the remainder will be ignored. Any orders arriving to the supplier when $n_{sX} \leq R$ will be ignored. Hence $n_{sX} \geq R$. The supplier may choose to set $R < 0$, in which case orders may arrive to the supplier even when the warehouse is empty. In this case $|R|$ is the number of unit requests that the supplier will "put on hold" until more units of the product arrive. If there are already $|R|$ unit requests being held then any order arriving to the supply node will be ignored, and if there are $|n_{sX}|$ ($R < n_{sX} < 0$) unit requests being held then only an additional $|R - n_{sX}|$ requests may be held. In either case ($R \geq 0$ or $R < 0$) $n_{sX} \geq R$.

When the state of the system is $\mathbf{n} = (n_{s0}, n_{s1} \ldots, n_{sX}, n_{dX})$ emission of class c customers from node i occurs at rate

$$\mu_{ic} \frac{\psi(\mathbf{n} - \mathbf{e}_{ic})}{\phi(\mathbf{n})},$$

where

$$\psi(\mathbf{n}) = \psi'(\mathbf{n})\phi(\mathbf{n}),$$

with

$$\phi(\mathbf{n}) = \prod_{\ell=1}^{n_{dX}} \xi_{dX}(\ell)^{-1} \prod_{c=0}^{X-1} \prod_{k=1}^{n_{sc}} \xi_{sc}(k)^{-1},$$

and $\psi'(\cdot)$ is set so that

$$\psi'(\mathbf{n}) = \begin{cases} 0, & \text{if } n_{sX} < R \text{ or } n_{ic} < 0 \text{ for } ic \in \{s0, s1, \ldots, s(X-1), dX\}, \\ 1, & \text{otherwise.} \end{cases} \tag{3.31}$$

This choice of functions allows state dependent emission rates of each class of customer at each node except class X customers at node s. The exception is made because we wish to store class X customers at node s until they are eliminated by signals arriving from node d. We accomplish this by setting $\mu_{sX} = 0$ so that node s does not emit any class X customers.

An order processed at node d *intends*, with probability $r(t)$, to consume/purchase t, $(t \geq 1)$, units of the finished product stored at node s. This is accomplished by routeing class X customers emitted from node d to node s as type t, $t \geq 1$, class X signals with probability $r(t)$. With this choice of $\psi(\cdot)$ and according to probabilities (3.4) and (3.5) a type t signal arriving to node s with the system in state $\mathbf{n}$ will:

- be ineffective with probability 1 if $n_{sX} \leq R$,
- become a type $n_{sX} - R$ signal with probability 1 if $R < n_{sX} < t + R$,
- remain a type t signal with probability 1 if $n_{sX} \geq t + R$.

We set all other network parameters mentioned in Section 3.1 to be zero. The equilibrium distribution of this system for a set of given parameters can then be examined to determine any congestions in the sequential manufacturing process, and the effectiveness of different manufacturing and distributing processes may be compared.

A very simple example of such a process occurs when $\{X = 0,\ R = 0,\ r(1) = 1\}$. Each order requests only one unit of the product and the product arrives to the system ready made. For this case the invariant measure is, from Theorem 3.2.1,

$$\pi(\mathbf{n}) = K \prod_{\ell=1}^{n_d} \xi_d(\ell)^{-1} x_s^{n_s} x_d^{n_d},$$

where, from equation (3.8),

$$0 = \mu p_s - x_s \mu_d x_d \quad \text{and} \quad \mu_d x_d = \mu p_d,$$

so that

$$x_d = \frac{\mu}{\mu_d} p_d \quad \text{and} \quad x_s = \frac{p_s}{p_d}.$$

The conditions under which the system has an equilibrium distribution are $x_i < 1$ so that we require

$$\mu_d > \mu p_d \text{ and } p_s < p_d.$$

The probability that an arriving order cannot be fulfilled (the probability that $n_s = 0$) is

$$1 - x_s = \frac{p_d - p_s}{p_d},$$

which is well defined whenever the conditions for a stable equilibrium are satisfied.

3.5 Predator-Prey Models

Consider an ecosystem of N species contained in a finite area in which each species has at most one predator, but may prey on a number of different species. Such ecosystems could be described as pyramidal food chains. We assume that each species has a fixed diet, so that a particular species is predator of a fixed subset of the N species in the ecosystem, and will prey on a particular species in that subset for a fixed proportion of the time that it is seeking food. Members of each species arrive to the ecosystem due to migration or birth, and depart due to migration or death. Death may occur due to starvation (there may not be enough of a particular species present in the ecosystem) or other natural causes. We assume that the collective departure rate and collective rate at which members of a particular species seek food depends on the population of that species, but that the collective arrival rate is a constant.

To model this we represent each species by a single node in an N node network, with an additional node (node 0) representing the outside of the ecosystem, and set

$$\psi(\mathbf{n}) = \psi'(\mathbf{n})\phi(\mathbf{n}),\ \forall \mathbf{n} \in S,$$

with

$$\phi(\mathbf{n}) = \prod_{i=1}^{N} \prod_{\ell=1}^{n_i} \xi_i(\ell)^{-1}.$$

The rate at which species $i \in \mathcal{L}$ enters the ecosystem is

$$\mu_0 p_i \frac{\psi(\mathbf{n})}{\phi(\mathbf{n})} = \mu_0 p_i \psi'(\mathbf{n}),$$

where

$$p_i = \mathrm{P}(\text{an arrival to the ecosystem is a member of species } i),$$

and

$$\sum_{i=1}^{N} p_i = 1.$$

The emission rate of species i individuals from node $i \in \mathcal{L}$ is

$$\mu_i \frac{\psi(\mathbf{n} - \mathbf{e}_i)}{\phi(\mathbf{n})} = \mu_i \psi'(\mathbf{n} - \mathbf{e}_i)\xi_i(n_i).$$

These emissions represent deaths or migrations if the emitted individual is transferred to node 0, but represent a member of species i hunting a member of species j if the emitted individual is transferred to node $j \in \mathcal{L}$. Once emitted from node i an individual will be transferred to node $j \in \mathcal{L} \cup \{0\}$ as a type 0 signal with probability $d^s_{ij}(0)$. The signal transfer probabilities fully describe the diets of each species in the ecosystem, and we require that

$$\sum_{j=0}^{N} d^s_{ij}(0) = 1 \text{ for } 1 \leq i \leq N.$$

Type 0 signals arriving to node 0 have no further effect on the network, so that for this example $d^s_{i0}(0)$ is equivalent to the probability d^+_{i0} of Section 3.1. The type 0 signal will be ineffective at node $j \in \mathcal{L}$ when the system is in state $\mathbf{n}$ with probability

$$1 - \frac{\psi(\mathbf{n} - \mathbf{e}_j)}{\psi(\mathbf{n})} = 1 - \frac{\psi'(\mathbf{n} - \mathbf{e}_j)}{\psi'(\mathbf{n})}\xi_j(n_j).$$

This is the probability that a member of species i, while hunting species j, will not find a member of species j to eat, and consequently die of starvation. Effective type 0 signals trigger a transfer of individuals from node j to node i with probability one (since species i is the only predator of species j) reducing the queue length at node j by one and increasing the queue length at node i by one. This represents a member of species i eliminating a member of species j from the ecosystem and returning to its base node.

The functions $\xi_i(\cdot)$, $i \in \mathcal{L}$, are used to model the dependence of the emission rates and the signal effectiveness probabilities on $\mathbf{n}$. A simple choice for $\xi_i(\cdot)$ would be

$$\xi_i(\ell) = \frac{\ell}{\ell + C},$$

where C is a positive constant. This would make the emission rate of species i proportional to n_i. $\psi'(\cdot)$ can again be chosen in any way we wish and, in particular, can be used, as in Section 3.3.1, to create lower boundaries for the state space. A simple choice for $\psi'(\cdot)$ would be

$$\psi'(\mathbf{n}) = \begin{cases} 0, & \text{if } n_i < 0 \text{ for any } i \in \mathcal{L}, \\ 1, & \text{otherwise.} \end{cases}$$

This would ensure that if $n_i = 0$ then additional members of species i could enter the ecosystem, but the emission rate from node i would be zero.

We may now determine the equilibrium distribution for the population levels in the ecosystem, or even construct optimal diets for each species (find optimal parameters $d^s_{ij}(0)$) to maximise population levels within the ecosystem.

A very simple example is an ecosystem of 3 species. Species 1 is the only predator of both species 2 and 3, and we are only interested in this level of the predator-prey interaction. We assume that ample food is available for species 2 and 3 so that they form the lowest level of the food chain in which we are interested. In this example we will only model departures from the system due to predator-prey interaction and due to starvation of predators. Let μ_0 be the total arrival rate into the system, p_i be the probability that an arrival to the system is a member of species i for $i =\in \{1, 2, 3\}$, μ_1 be the rate that members of species 1 seek food, and set $\{\mu_2 = \mu_3 = 0,\ d^s_{12}(0) = r,\ d^s_{13}(0) = 1 - r\}$, so that the diet of species 1 is fully described by the parameter r. The functions $\psi(\cdot)$, $\phi(\cdot)$, and $\xi_i(\cdot)$ are as suggested above. All other parameters are set to zero. Thus the total rate at which species 1 searches for food is proportional to its population, and a member of species 1 will hunt a member of species 2 and 3 with probabilities r and $(1 - r)$ respectively. We will investigate any restrictions on r which are necessary for stability of the system. The invariant measure for the system is given by equation (3.7) where, by equation (3.8),

$$\begin{aligned}\mu_1 x_1 &= \mu_0 p_1 + \mu_1 x_1(x_2 r + x_3(1-r)),\\ 0 &= \mu_0 p_2 - x_2 \mu_1 x_1 r,\\ 0 &= \mu_0 p_3 - x_3 \mu_1 x_1 (1-r).\end{aligned}$$

Thus

$$x_1 = \frac{\mu_0}{\mu_1}, \; x_2 = \frac{p_2}{r}, \; \text{and } x_3 = \frac{p_3}{(1-r)}.$$

For stability in the system we require that $x_i < 1$ which constrains the choice of r to $r \in (p_2, 1-p_3)$. If these constraints are not satisfied then the population of either species 2 or species 3 will explode. $x_1 < 1$ also requires that $\mu_0 < \mu_1$ to ensure that the population of species 1 is stable.

3.6 Networks of Customer and Resource Queues

Consider a network of N queues partitioned so that queues labelled $1, 2, \ldots, N_R$ are "resource" queues and $N_R + 1, \ldots, N$ are "customer" queues. Resources arrive to and are stored in the "resources" partition. The customers are processed in the "customer" partition until they require access to some resources. They accomplish this by being routed to the "resources" partition as a trigger. We consider an additional node, node 0, to represent the outside of the system.

It is assumed that customers are never signalled, customers never become resources, resources are never served but may be triggered, and that triggers cannot arrive from outside the network. That is, simplifying notation by setting $d^s_{ij}(0) = d^t_{ij}$, we assume

$$\begin{aligned}d^t_{0j} &= 0, \quad j \in \{1, 2, \ldots, N\},\\ d^t_{ij} &= 0, \quad i \in \{1, 2, \ldots, N\}, \quad j \in \{N_R + 1, \ldots, N\},\\ d^+_{ij} &= 0, \quad i \in \{N_R + 1, \ldots, N\}, \quad j \in \{1, 2, \ldots, N_R\},\\ \mu_i &= 0, \quad i \in \{1, 2, \ldots, N_R\}.\end{aligned}$$

We do not include type $t \geq 1$ signals in these networks, and so

$$\sum_{j=1}^{N} \left[d_{ij}^{+} + d_{ij}^{t}\right] = 1, \ \forall i \in \{0, N_R + 1, \ldots, N\}.$$

Since the resources are never served it is not necessary to define d_{ij}^{t} and d_{ij}^{+} for $i \in \{1, 2, \ldots, N_R\}$. However we are still free to make the choice

$$\phi(\mathbf{n}) = \psi(\mathbf{n}) = \prod_{i=1}^{N} \prod_{\ell=1}^{n_i} \nu_i(\ell)^{-1},$$

where

$$\nu_j(n_j) = \begin{cases} 1, & n_j > 0, \\ 0, & n_j = 0, \end{cases} \quad \text{for } j \in \{1, \ldots, N_R\},$$

and assume that $\nu_i(n_i)$ is arbitrary for $i \in \{N_R + 1, \ldots, N\}$.

Customers are being served at customer queues at state dependent rates and can move between customer queues as they would in a standard Jackson network. However, a customer may become a trigger and will do so with intensity $\mu_i \nu_i(n_i) d_{ij}^{t} \nu_j(n_j)$ for $i \in \{N_R + 1, \ldots, N\}$ and $j \in \{1, 2, \ldots, N_R\}$, i.e.

$$\begin{cases} \mu_i \nu_i(n_i) d_{ij}^{t}, & \text{if } n_j \geq 0, \\ 0, & \text{if } n_j < 0. \end{cases}$$

Thus when a resource is available a customer can be served from customer queue i at rate $\mu_i \nu_i(n_i)$, and trigger a resource at queue j with probability d_{ij}^{t}. If there is no resource available at queue j the customer is served from queue i and leaves the network.

Customers are never triggered so we have no need to define q_{ij} for $i \in \{N_R + 1, \ldots, N\}$. If we place no restrictions on q_{ij} for $i \in \{1, 2, \ldots, N_R\}$ then a customer who triggers a resource can leave the network with the resource ($j = 0$), leave the network but move the resource to another resource queue ($j \in \{1, 2, \ldots, N_R\}$) or remove the resource from the network and return to a customer queue ($j \in \{N_R + 1, \ldots, N\}$).

$$\sum_{j=0}^{N} q_{ij} = 1, \ \forall i \in \{1, 2, \ldots, N_R\}.$$

The equilibrium distribution for the network is, from Theorem 3.2.1,

$$\pi(\mathbf{n}) = K \left[\prod_{i=1}^{N_R} x_i^{n_i}\right] \left[\prod_{i=N_R+1}^{N} x_i^{n_i} \prod_{\ell=1}^{n_i} [\nu_i(\ell)]^{-1}\right].$$

In this case the traffic equations for the x_i's simplify to

$$\mu_i x_i = \mu_0 d^{+}_{0i} + \sum_{h=N_R+1}^{N} \mu_h x_h \left[d^{+}_{hi} + \sum_{k=1}^{N_R} x_k d^{t}_{hk} q_{ki} \right], \quad i \in \{N_R+1, \ldots, N\},$$

and

$$x_i \sum_{j=N_R+1}^{N} \mu_j x_j d^{t}_{ji} = \mu_0 d^{+}_{0i} + \sum_{h=N_R+1}^{N} \mu_h x_h \sum_{k=1}^{N} x_k d^{t}_{hk} q_{ki}, \quad i \in \{1, 2, \ldots, N_R\}.$$

3.7 Communications Protocols

3.7.1 Once Only Verification Protocol

The examples presented here are approximations to simple verify and retransmit protocols in packet switched networks.

Consider a network of 3 nodes, as illustrated in Figure 3.1, in which original packets arrive to, and are queued at, node A awaiting transmission to node C. After transmission the packets are queued at node C awaiting error checking. Duplicate packets, to those arriving at node A, arrive at node B in an independent Poisson stream. This is where the approximation is made - in reality arrivals to nodes A and B occur simultaneously. However, as in the Erlang fixed point approximation, such assumptions can produce accurate results especially when arrival rates are large.

Following the error check on an original packet at node C a message is sent to node B indicating the error status of the original packet. With probability $1-\varepsilon$ the packet is found to be error free at node C and a trigger leaves node C to delete the duplicate packet at node B. With probability ε the packet is found to contain errors at node C and returns to node B as a trigger to transfer the duplicate packet from node B to node A for retransmission. Now the duplicate packet is queueing for transmission from node A to node C. When a duplicate packet is served at node C it always leaves the network either because it is error free, and therefore usable, or it still contains an error and is dropped.

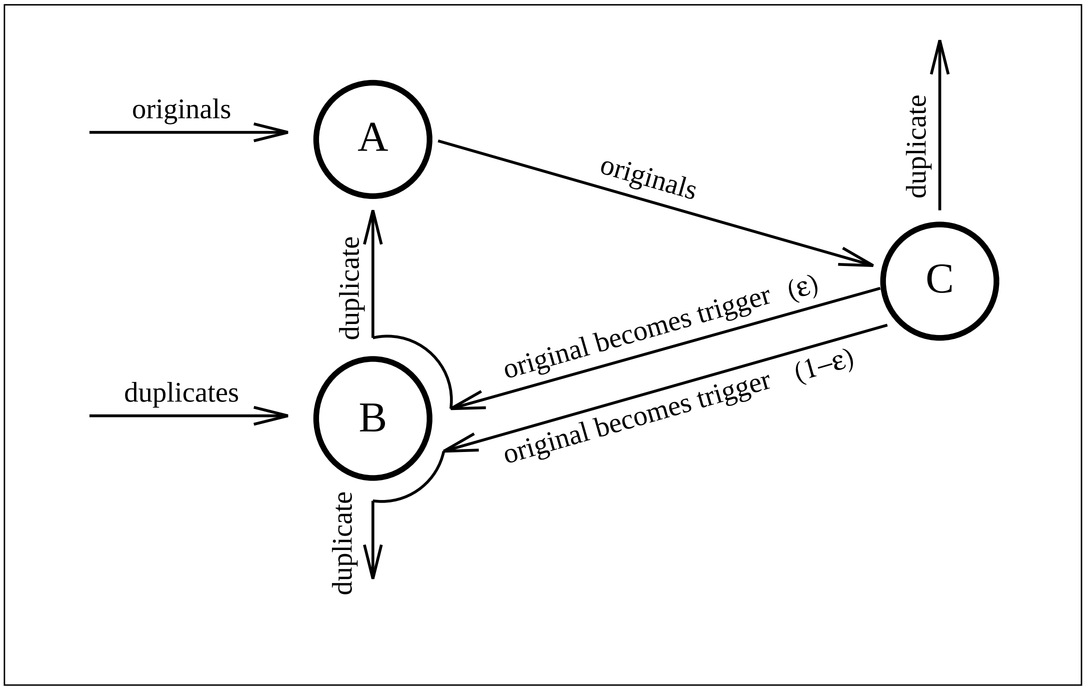

Figure 3.1: Once only verification protocol

The equilibrium distribution for this network has a product form solution, extended to account for the different customer classes (original and duplicate) and is modelled using the ideas of Section 3.6 by treating queues A and C as customer queues, with customers (packets) from queue A always routed as positive customers to queue C. Queue B is a resource queue which loses resources only by being triggered. The triggering always arrives from queue C when original packets are served there and found to contain errors. The service rates at queues A and C can be arbitrary "Jackson type". A trigger arriving to queue B sends a packet to queue A with probability ε for retransmission and out of the network with probability $1 - \varepsilon$.

In this particular example we could have returned a proportion ε of the original packets directly to node A for retransmission, changing their classes to "duplicate" in the process and never use node B or triggering. This would remove the approximation made in the model. However, the advantage of retaining node B is in being able to observe directly the buffer size required to hold duplicate messages.

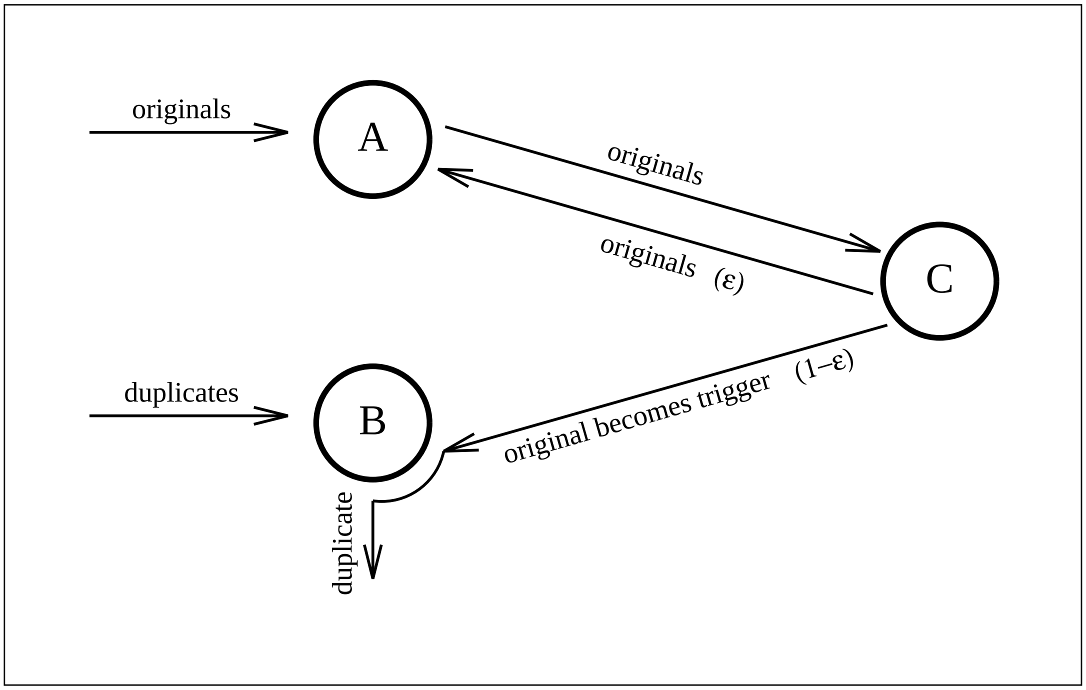

Figure 3.2: Repeated verification protocol

3.7.2 Repeated Verification Protocols

In this example, if an error is located in a transmitted packet the packet is retransmitted until it is correct. It is assumed that there is a constant probability ε that any packet arriving to node C (see Figure 3.2) will contain an error and therefore need to be retransmitted irrespective of whether it is an original or a duplicate. This protocol can be modelled as shown in Figure 3.2 with packets leaving node C returned to node A with probability ε representing the need for a retransmission. Note that when this occurs the content of the buffer, at node B, is unchanged because the duplicate packet must be retained until the packet is transmitted successfully. A duplicate is eventually removed from the buffer when a packet is served at node C and, with probability $1-\varepsilon$, becomes a trigger (and in fact a negative customer) is routed to node B and removes one packet. Again note that the model is one of the class described in Section 3.6, this time with only a single customer class in the network. It therefore follows that the service rates at nodes A and C can be a function of the number of packets at the nodes.

If, on the other hand, a packet can be retransmitted at most N times this can be achieved using these models by reintroducing classes of customers into the model of Figure 3.2 so that each time a packet is returned to node A from node C it changes class from i to $i+1$. When it reaches class N it behaves as if a packet in Figure 3.1 moving from node C to node B as a trigger, releasing the duplicate and taking it out of the network with probability $1-\varepsilon$ or sending it to node A for its final retransmission with probability ε.

Chapter 4

Quasi-reversibility

Each of the queueing networks presented in this thesis have product form invariant measures. A network, formed by coupling/linking quasi-reversible processes together to form a quasi-reversible network, also has a product form invariant measure. In this chapter we investigate the applicability of previous results on quasi-reversibility to the networks discussed in this thesis, and demonstrate a new method of coupling to allow transitions to affect more than two different nodes simultaneously. This is necessary if we are to model the type 0 signals (triggers) of Chapter 3 by linking quasi-reversible nodes.

Quasi-reversibility is a property of some processes, first identified by Muntz [58], which is closely related to partial balance and insensitivity. Quasi-reversible nodes may be coupled to form a network, which is itself quasi-reversible, and whose invariant measure is a product of the invariant measures of the nodes. Kelly [53] comments on the relationship between partial balance and quasi-reversibility, indicating that partial balance is concerned with customer behaviour, and that quasi-reversibility focusses attention on the nodes of the network.

Basket, Chandy, Muntz and Palacios [3] obtain the so-called BCMP network by connecting nodes which individually satisfy partial balance. The derivation could have been made by showing that each type of node allowed in a BCMP network is quasi-reversible. Pollett [60] indicates that BCMP networks include nodes that, in isolation, may not be reversible, but are quasi-reversible.

Networks of quasi-reversible processes, obtained using a number of coupling techniques, have been studied by Walrand and Varaiya [71], Kelly [52] [53], Pollett [60] [61], Whittle [76], Henderson, Pearce, Pollett and Taylor [39] and Henderson, Pearce, Schassberger and Taylor [44]. Different versions of the formulation of quasi-reversibility have been presented in each of these papers to highlight one feature or another of the particular networks being discussed. We present here a modified version of the formulation in [44], which is itself a modified version of [39] and earlier papers.

4.1 Quasi-reversible Processes

Consider a regular, stationary Markov process defined on a countable set of states S with transition rates $q(n,n')$, for $n, n' \in S$, and positive invariant measure $\{m(n),\ n \in S\}$, which must satisfy

$$m(n)q(n) = \sum_{n' \in S} m(n')q(n',n), \tag{4.1}$$

where $q(n) = \sum_{n' \in S} q(n,n')$. In this chapter we will allow transitions to occur that do not cause a change of state, so that it is possible that $q(n,n) > 0$ for some $n \in S$. This is done for notational convenience and does not affect the invariant measure as the term $m(n)q(n,n)$ will appear on both sides of equation (4.1). The set of all possible transitions is

$$\mathcal{Q} = \{(n,n') : q(n,n') > 0\}.$$

Define

$$q'(n,n') = \frac{m(n')q(n',n)}{m(n)}, \quad \forall (n,n') \text{ such that } (n',n) \in \mathcal{Q}, \tag{4.2}$$

and $q'(n) = \sum_{n' \in S} q'(n,n')$. The rates $q'(\cdot,\cdot)$ are the transition rates of the process when it is observed in "reverse time".

The following lemma is presented as Theorem 1.13 in Kelly [52].

Lemma 4.1.1 *If we can find a collection of transition rates,* $\mathcal{Q}' = \{q'(n,n'),\ n,n' \in S\}$*, together with a collection of positive numbers,* $\mathbf{m} = \{m(n),\ n \in S\}$*, which satisfy*

$$m(n)q'(n,n') = m(n')q(n',n), \quad n,n' \in S, \tag{4.3}$$

then $\mathbf{m}$ *is an invariant measure for* $\mathcal{Q}$ *if and only if* $q'(n) = q(n),\ \forall n \in S$*.*

Define a collection of pairs of probabilities $\mathcal{F} = \{f^a(t;n,n'), f^d(t;n,n')\}$ indexed by a set of transition types $\mathcal{T}$ and the transitions $\mathcal{Q}$, to be, for $t \in \mathcal{T}$ and $(n,n') \in \mathcal{Q}$,

$$f^a(t;n,n') = P\{\text{transition } n \to n' \text{ is a class } t \text{ ``arrival''}\},$$

$$f^d(t;n,n') = P\{\text{transition } n \to n' \text{ is a class } t \text{ ``departure''}\}.$$

Type $t \in \mathcal{T}$ events ("arrivals" and "departures") need not refer to the physical arrivals and departures of customers in a process with multiple customer classes (see Chapter 3), but may instead be arbitrarily assigned to refer to any entity whose movement defines a transition. For example, in this chapter we consider the deposit of a batch of k negative customers at a node to be a type $-k$ "arrival". There may be some transitions that occur in the Markov process which are not associated with any of the transition types in $\mathcal{T}$. That is, we may not be able to fully describe each possible transition of the process using the set of labels $\mathcal{T}$. We do not allow events of type $t \in \mathcal{T}$ and $t' \in \mathcal{T}$ to occur simultaneously, but events $t \in \mathcal{T}$ and $t' \notin \mathcal{T}$ may do so. Thus, for all $(n,n') \in \mathcal{Q}$,

$$\sum_{t \in \mathcal{T}} \left[f^a(t;n,n') + f^d(t;n,n') \right] \le 1.$$

For $(t,n) \in \mathcal{T} \times S$ the total rate of type t "arrivals" in state n is

$$\alpha(t,n) = \sum_{n' \in S} q(n,n') f^a(t;n,n'). \tag{4.4}$$

Similarly, equation (4.3) and the invariant measure $m(\cdot)$, implies, for $(t,n) \in \mathcal{T} \times S$, that

$$\begin{aligned} \beta(t,n) &= \sum_{n' \in S} \frac{m(n')}{m(n)} q(n',n) f^d(t;n',n), \\ &= \sum_{n' \in S} q'(n,n') f^d(t;n',n), \end{aligned} \tag{4.5}$$

is the total rate of type t "departures" which leave the process in state n. Alternatively, equation (4.5) may be identified as the total rate of "arrivals" of type t when the process is in state n and the network is observed in "reverse time".

Definition 4.1.1 *A process is quasi-reversible with respect to* $\{\mathcal{T}, \mathcal{F}\}$ *if there exists* $\{\alpha(t),\ \beta(t),\ \textit{for}\ t \in \mathcal{T}\}$ *such that*

$$\alpha(t,n) = \alpha(t) \text{ and } \beta(t,n) = \beta(t),\ \forall (t,n) \in \mathcal{T} \times S.$$

Kelly [53] describes the consequences of quasi-reversibility to be the following. If a stationary Markov process is quasi-reversible, and its positive invariant measure can be normalised to give the equilibrium distribution, then the state of the process at time τ is independent of:

- the stream of type t "arrivals" subsequent to time τ,
- the stream of type t "departures" prior to time τ.

This implies that type t "arrivals" and type t "departures" occur as independent Poisson streams.

Definition 4.1.2 *A quasi-reversible Markov process is internally balanced if*

$$\sum_{t \in \mathcal{T}} \alpha(t) = \sum_{t \in \mathcal{T}} \beta(t). \tag{4.6}$$

4.2 Networks of Quasi-reversible Nodes

Quasi-reversible nodes have traditionally been coupled together to form a network of N nodes, labelled from a set $\mathcal{L} = \{1, 2, \ldots, N\}$, by allowing a type $t \in \mathcal{T}$ "departure" at node $i \in \mathcal{L}$ to trigger a type $t' \in \mathcal{T}$ "arrival" at node $j \in \mathcal{L}$ with probability $p_{ij}(t, t')$, and by allowing exogenous "arrivals" of type t to node i to occur at rate $\nu_i(t)$. The resulting network will be quasi-reversible if the parameters

$$\{p_{ij}(t,t'),\ \nu_i(t),\ i,j \in \mathcal{L},\ t,t' \in \mathcal{T}\}$$

provide the network with some degree of balance. The balance required is reflected in a set of traffic equations which must hold for the network to have a product form invariant measure. These traffic equations vary between networks. For example, a Jackson network has traffic equations given by equations (1.5), but a network with single positive and negative customers has traffic equations given by equations (1.9).

Henderson, Pearce, Pollett and Taylor [39] show how to couple N quasi-reversible nodes to obtain a quasi-reversible network. Their method of coupling, which we present here, is a generalisation of the method adopted by Kelly [53]. In isolation, let the process for the queue length at node $i \in \mathcal{L}$ be characterised by the state space S_i and transition rates $\{q_i(n,n') : (n,n') \in \mathcal{Q}_i\}$, and be quasi-reversible with respect to $\{\mathcal{T}, \mathcal{F}_i\}$. Each one of the N nodes is quasi-reversible with respect to the common set of transition types $\mathcal{T}$. The state space of the coupled process is $S \subseteq S_1 \times S_2 \times \ldots \times S_N$ and a typical state is $\mathbf{n} = (n_1, n_2, \ldots, n_N)$, so that $n_i \in S_i$. Denote by $\mathbf{n} + (n_i' - n_i)\mathbf{e}_i$ the state which is the same as $\mathbf{n}$ except that the i-th process has changed state from n_i to n_i'. We set the transition rates $q(\cdot,\cdot)$ for the coupled process to be, for $i,j \in \mathcal{L}$, $n_i \neq n_i',\ n_i, n_i' \in S_i,\ (n_i,n_i') \in \mathcal{Q}_i$, and $n_j \neq n_j',\ n_j, n_j' \in S_j,\ (n_j,n_j') \in \mathcal{Q}_j$,

$$\begin{aligned} q(\mathbf{n}, \mathbf{n} + (n_i' - n_i)\mathbf{e}_i + (n_j' - n_j)\mathbf{e}_j) \\ = \sum_{t \in T} \sum_{t' \in T} q_i(n_i, n_i') f_i^d(t; n_i, n_i') p_{ij}(t,t') \frac{q_j(n_j, n_j') f_j^a(t'; n_j, n_j')}{\alpha_j(t')} \\ + \sum_{t \in T} \sum_{t' \in T} q_j(n_j, n_j') f_j^d(t; n_j, n_j') p_{ji}(t,t') \frac{q_i(n_i, n_i') f_i^a(t'; n_i, n_i')}{\alpha_i(t')}, \end{aligned} \tag{4.7}$$

and for $i \in \mathcal{L},\ n_i \neq n_i',\ n_i, n_i' \in S_i,\ (n_i, n_i') \in \mathcal{Q}_i$,

$$\begin{aligned} q(\mathbf{n}, \mathbf{n} + (n_i' - n_i)\mathbf{e}_i) \\ = \sum_{j=1}^{N} \sum_{t \in T} \sum_{t' \in T} q_j(n_j, n_j) f^d(t; n_j, n_j) p_{ji}(t,t') \frac{q_i(n_i, n_i') f_i^a(t'; n_i, n_i')}{\alpha_i(t')} \end{aligned}$$

$$+\sum_{t\in T} q_i(n_i, n_i') f_i^d(t; n_i, n_i') p_{i0}(t)$$
$$+\sum_{t\in T} \frac{\nu_i(t) q_i(n_i, n_i') f_i^a(t; n_i, n_i')}{\alpha_i(t)}$$
$$+q_i(n_i, n_i')\left[1 - \sum_{t\in T}\left[f^a(t; n_i, n_i') + f^d(t; n_i, n_i')\right]\right], \quad (4.8)$$

where

$$p_{i0}(t) = 1 - \sum_{j=1}^{N} \sum_{t'\in \mathcal{T}} p_{ij}(t, t'),\ i \in \mathcal{L}.$$

Lemma 4.2.1 *If* $\{\alpha_i(t),\ \beta_i(t),\ \forall i \in \mathcal{L}, (t, n) \in \mathcal{T} \times S\}$ *satisfy the equation*

$$\alpha_i(t) = \nu_i(t) + \sum_{j=1}^{N} \sum_{t'\in \mathcal{T}} \beta_j(t') p_{ji}(t', t),\ i \in \mathcal{L}, \quad (4.9)$$

then the coupled process has invariant measure

$$m(\mathbf{n}) = \prod_{i=1}^{N} m_i(n_i), \quad (4.10)$$

and the process itself is quasi-reversible, where $m_i(\cdot)$ *is the invariant measure for the i-th process.*

The proof for this Lemma may be found in [39], in which the reverse time transition rates $q'(\mathbf{n}', \mathbf{n})$ are defined to ensure that $m(\mathbf{n})q(\mathbf{n}, \mathbf{n}') = m(\mathbf{n}')q'(\mathbf{n}', \mathbf{n})$, and it is shown that $q(\mathbf{n}) = q'(\mathbf{n})$ so that the network satisfies the conditions of Lemma 4.1.1. The following Corollary is given in [44].

Corollary 4.2.1 *Fix the numbers* $\alpha_i(t)$, $\beta_i(t)$, $\nu_i(t)$ *and* $p_{ij}(t, t')$, *for* $i, j \in \mathcal{L}$. *Then for the network, formed by coupling* N *quasi-reversible processes with arrival and departure rates* $\alpha_i(t)$ *and* $\beta_i(t)$ *using equations (4.7) and (4.8), to have an invariant measure given by equation (4.10), it is necessary and sufficient that the traffic equations (4.9) be satisfied.*

4.3 Queueing Networks with Batches of Negative Customers

Gelenbe [27], [28] states that queueing networks with positive and negative customers are not quasi-reversible in the usual sense. This is true if we restrict the entities, indexed by the set $\mathcal{T}$, to be physical (positive) customers only, but the definition of quasi-reversibility allows for more generality. Henderson, Pearce, Schassberger and Taylor [44] show that the queueing networks of Chapter 2 have a product form equilibrium distribution by showing that they are quasi-reversible with respect to a set of transition types which do not refer only to positive customers. We repeat their results here for completeness.

Consider, in isolation, a single node of the queueing network described in Chapter 2. The state space of the underlying Markov process for the queue length of the node is Z_+. Assume that the total arrival rate of positive customers and batches of k negative customers, for $k = 1, 2 \ldots, K$ (with K possibly infinite), is $\lambda > 0$ and γ_k respectively, and that the service rate of the node is $\mu > 0$. We require $\sum_{k=1}^{K} \gamma_k < \infty$. The process evolves using the following transition rates:

$$
\begin{aligned}
q(n, n+1) &= \lambda, \\
q(n, n-1) &= \begin{cases} \mu + \gamma_1, & \text{for } n > 1, \\ \mu + \sum_{k=1}^{K} \gamma_k, & \text{for } n = 1, \end{cases} \\
q(n, n-k) &= \gamma_k, \quad \text{for } n > k \text{ and } 1 < k \leq K, \\
q(n, 0) &= \sum_{k=n}^{K} \gamma_k, \quad \text{for } 1 < n \leq K, \\
q(0, 0) &= \sum_{k=1}^{K} \gamma_k,
\end{aligned}
\tag{4.11}
$$

with $q(n, n')$ zero otherwise. An invariant measure, $m(n)$, for the process must satisfy

$$
m(n)\lambda = m(n+1)\mu + \sum_{k=1}^{K} m(n+k) \sum_{\ell=k}^{K} \gamma_\ell,
$$

as they are the partial balance equations obtained when the state space is cut after state n, the solution of which is $m(n) = z_0^n$, where z_0 is the unique positive root of

$$\lambda = z\mu + \sum_{k=1}^{K} z^k \sum_{\ell=k}^{K} \gamma_\ell. \tag{4.12}$$

Now define the probabilities $f^a, f^d \in \mathcal{F}$ to be:

$$\begin{aligned}
f^a(+1; n, n+1) &= 1, \\
f^d(+1; n, n-1) &= \begin{cases} \dfrac{\mu}{\mu+\gamma_1}, & \text{for } n > 1, \\ \dfrac{\mu}{\mu + \sum_{k=1}^{K} \gamma_k}, & \text{for } n = 1, \end{cases} \\
f^a(-1; n, n-1) &= \frac{\gamma_1}{\mu+\gamma_1}, \quad \text{for} \quad n > 1, \\
f^a(-k; n, n-k) &= 1, \quad \text{for } n > k \text{ and } 1 < k \leq K, \\
f^a(-k; n, 0) &= \begin{cases} \dfrac{\gamma_k}{\sum_{\ell=n}^{K} \gamma_\ell}, & \text{for } (n=0) \text{ or } (n \leq k \text{ and } 1 < k \leq K), \\ \dfrac{\gamma_k}{\mu + \sum_{\ell=1}^{K} \gamma_\ell}, & \text{for } n = 1, \end{cases}
\end{aligned} \tag{4.13}$$

with $f^a(t; n, n')$ and $f^d(t; n, n')$ zero otherwise. The index set for the transitions of the network is $\mathcal{T} = \{+1, -1, -2, \ldots, -K\}$. A type $-k$, $1 \leq k \leq K$, event corresponds to the arrival of a batch of k negative customers, and a type $+1$ event is either an arrival or departure, due to service completion, of a positive customer. There are no type $-k$ departures in this system. The process will be quasi-reversible with respect to $\mathcal{T}$ and $\mathcal{F}$ if we can show, from (4.4) and (4.5), that the "arrival" and "departure" streams of type $t \in \mathcal{T}$ entities are state independent. This is clearly the case for the "arrival" streams since

$$\alpha(+1, n) = \lambda, \;\; \alpha(-k, n) = \gamma_k, \;\; \forall n \in S, \;\; 1 \leq k \leq K. \tag{4.14}$$

Trivially

$$\beta(-k, n) = 0, \;\; \forall n \in S, \;\; 1 \leq k \leq K,$$

which is state independent, and so we only have to show that the "departure" stream of type $+1$ entities is state independent.

$$\begin{aligned}
\beta(+1, n) &= q'(n, n+1) f^d(+1; n+1, n), \\
&= \frac{m(n+1) q(n+1, n)}{m(n)} f^d(+1; n+1, n), \\
&= z_0 \mu,
\end{aligned} \tag{4.15}$$

which is independent of state n. Thus the process is quasi-reversible. Equation (4.6) must hold for this network to be internally balanced, that is

$$\lambda + \sum_{k=1}^{K} \gamma_k = z_0 \mu,$$

which is clearly not the case in this situation. Other authors refer to systems which are internally balanced as "customer preserving", because, in every state, the total rate of "arrivals" is balanced by the total rate of "departures". It follows then, that this system, in which customers may be destroyed, is not internally balanced. The throughput of positive customers at the single server is the rate of service completions or, alternatively, the rate of type +1 departures, which is $\beta(+1) = z_0\mu$. Thus equation (4.12) becomes

$$\alpha(+1) = \beta(+1) + \sum_{k=1}^{K} z_0^k \sum_{\ell=k}^{K} \alpha(-\ell).$$

We can couple N of these nodes together to form the queueing network described in Chapter 2. The following theorem, presented in [44], confirms the results of Chapter 2 and follows directly from Lemma 4.2.1.

Theorem 4.3.1 *An invariant measure for the coupled network of nodes with positive and negative customer arrivals is given by*

$$m(\mathbf{n}) = \prod_{i=1}^{N} z_i^{n_i},$$

where z_i, $i \in \mathcal{L}$, is the unique positive root of the equation

$$\alpha_i(+1) = \beta_i(+1) + \sum_{k=1}^{K} z^k \sum_{\ell=k}^{K} \alpha_i(-\ell), \;\; i \in \mathcal{L}, \tag{4.16}$$

and for $t \in \mathcal{T}$

$$\alpha_i(t) = \nu_i(t) + \sum_{j=1}^{N} \beta_j(+1) p_{ji}(+1, t), \;\; i \in \mathcal{L}. \tag{4.17}$$

Proof: The theorem is easily substantiated for the case $K = 1$, (in which case the network is equivalent to those of Gelenbe [27]), by replacing the coupled network parameters with their Gelenbe network equivalents according to Table 4.1.

Coupled Network	Gelenbe Network
$\alpha_i(+1)$	$\lambda^+(i)$
$\alpha_i(-1)$	$\lambda^-(i)$
$\nu_i(+1)$	$\Lambda(i)$
$\nu_i(-1)$	$\lambda(i)$
μ_i	$r(i)$
$\beta_i(+1)$	$q_i r(i)$
$p_{ij}(+1,+1)$	$p^+(i,j)$
$p_{ij}(+1,-1)$	$p^-(i,j)$

Table 4.1: Parameter equivalence between coupled and Gelenbe networks

Following the substitutions, equations (4.16) and (4.17) are equivalent to equations (2.27) and (2.29).

For the case of batch arrivals of negative customers we let $K \to \infty$ and make the substitutions given in Table 4.2. Equations (4.17) become

Coupled Network	Batch Negative Network
$\nu_i(+1)$	$Y_0 d^+_{0i}$
$\nu_i(-k),\ k \geq 1$	$Y_0 d^-_{0i}(k)$
$\beta_i(+1)$	Y_i
z_i	$\left(\frac{Y_i}{\mu_i}\right)$
$p_{ij}(+1,+1)$	d^+_{ij}
$p_{ij}(+1,-k)\ k \geq 1$	$d^-_{ij}(k)$

Table 4.2: Parameter equivalence between coupled and batch negative networks

$$\alpha_i(+1) = \sum_{j=0}^{N} Y_j d^+_{ji},$$

and

$$\alpha_i(-k) = \sum_{j=0}^{N} Y_j d^-_{ji}(k),$$

so that equation (4.16) becomes

$$\sum_{j=0}^{N} Y_j d^+_{ji} = Y_i + \sum_{k=1}^{\infty} \left(\frac{Y_i}{\mu_i}\right)^k \sum_{\ell=k}^{\infty} \sum_{j=0}^{N} Y_j d^-_{ji}(k),$$

which is equivalent to the traffic equations of Chapter 2, namely equations (2.9).

□

The networks which have been discussed in this section have a state space Z_+^N, restricting the queue length of customers at each node to be non-negative. If, however, we allow queue lengths to become negative, as in Chapter 3, then the transition rates for the network are simplified to

$$\begin{aligned} q(n, n+1) &= \lambda, \\ q(n, n-1) &= \mu + \gamma_1, \\ q(n, n-k) &= \gamma_k, \ 1 < k \leq K, \end{aligned}$$

with $q(n, n')$ zero otherwise. The probabilities $f^a, f^d \in \mathcal{F}$ are simplified in a similar way, and the network is quasi-reversible with

$$\alpha(+1, n) = \lambda, \ \alpha(-k, n) = \gamma_k, \ \beta(+1, n) = z_0\mu, \ \forall n \in S, \ 1 \leq k \leq K.$$

Clearly then, it is of more interest (and more difficult) to discuss quasi-reversibility when we restrict attention to queueing networks with non-negative queue lengths. Note that if there is no lower bound on the state space of the network with negative queue lengths, then the geometric invariant measure is not summable, and it cannot be normalised to obtain the equilibrium distribution.

4.4 Triggered Linking of Quasi-reversible Nodes

The queueing networks of Chapter 3 include type $k \geq 1$ signals (batches of k negative customers) and also type 0 signals (triggers). We have already shown that queueing networks, restricted to include only type $k \geq 1$ signals, are quasi-reversible. We will restrict attention, for the remainder of this section, to queueing networks, with triggers being the only signalling entities, and non-negative queue lengths of customers (being the more difficult case), and show that these networks are quasi-reversible. The networks under consideration include positive customers and triggers only, as described by Gelenbe [29], or as in Chapter 3 without state dependence and with $d^s_{ij}(k) = 0$ if $k \neq 0$. We do not allow triggered customers to be routed directly to the outside of the network, as these transitions are duplicated by the arrival of negative customers. Gelenbe [29] shows that these networks have a geometric product form equilibrium distribution.

The coupling of quasi-reversible nodes, described by equations (4.7) and (4.8), involves a change of state for at most two nodes of the network in any given transition. A triggering transition causes a simultaneous change of state for three nodes. Therefore, we need to be able to link three nodes together in one transition to obtain a network with triggered customer movement. To our knowledge this type of linking has not been achieved in a quasi-reversible

In the following, we could include transitions which correspond to the arrival of batches of negative customers, and derive a quasi-reversible network. We do not however, as this would complicate the analysis, and draw attention away from the emphasis of this section, which is the triggered linking of nodes.

Consider an isolated single server node with single positive customer arrivals. The transition rates $\{q(n, n'),\ (n, n') \in \mathcal{Q}\}$, and probabilities $\mathcal{F}$ for the underlying Markov process are

$$\begin{aligned} q(n, n+1) &= \lambda, \\ q(n, n-1) &= \mu,\ \text{for } n \geq 1, \\ f^a(+1; n, n+1) &= 1, \\ f^d(+1; n, n-1) &= 1,\ \text{for } n \geq 1, \end{aligned} \tag{4.18}$$

and the node is quasi-reversible with $\alpha(+1) = \lambda$ and $\beta(+1) = z\mu$, given that an invariant measure for the process is z^n. We will build a network with triggered customer movement by linking N of these nodes. The method of linking that we use to obtain the network is implicitly defined when we set the transition rates of the network. Let

$\mathcal{E}(\mathbf{n}) \subseteq \mathcal{L}$ and $\mathcal{E}(\mathbf{n})^c = \mathcal{L}\backslash\mathcal{E}(\mathbf{n})$ be the sets of nodes which are empty and non-empty in state $\mathbf{n}$ respectively, then we set the transition rates for a network with triggered linking to be

$$\begin{aligned}
q(\mathbf{n}, \mathbf{n} - \mathbf{e}_i - \mathbf{e}_j + \mathbf{e}_k) &= \mu_i p^s_{ij} q_{jk}, \quad i, j \in \mathcal{E}(\mathbf{n})^c,\ k \in \mathcal{L}, \\
q(\mathbf{n}, \mathbf{n} - \mathbf{e}_i + \mathbf{e}_j) &= \mu_i p^+_{ij} + \nu^s_i q_{ij}, \quad i \in \mathcal{E}(\mathbf{n})^c,\ j \in \mathcal{L}, \\
q(\mathbf{n}, \mathbf{n} - \mathbf{e}_i) &= \mu_i p^+_{i0} + \mu_i \sum_{j \in \mathcal{E}(\mathbf{n})} p^s_{ij}, \quad i \in \mathcal{E}(\mathbf{n})^c, \\
q(\mathbf{n}, \mathbf{n} + \mathbf{e}_i) &= \nu^+_i, \quad i \in \mathcal{L}, \\
q(\mathbf{n}, \mathbf{n}) &= \sum_{i \in \mathcal{E}(\mathbf{n})} \nu^s_i,
\end{aligned} \tag{4.19}$$

where ν^+_i and ν^s_i are the exogenous arrival rates of positive customers and triggers respectively to node i, and p^+_{ij}, p^s_{ij} and q_{ij} are the routeing probabilities of positive customers, triggers and triggered customers respectively. We require that

$$\sum_{j=1}^{N} (p^+_{ij} + p^s_{ij}) + p^+_{i0} = 1, \quad \text{and} \quad \sum_{j=1}^{N} q_{ij} = 1, \ \text{for } i \in \mathcal{L}.$$

Thus

$$q(\mathbf{n}) = \sum_{i=1}^{N} (\nu^+_i + \nu^s_i) + \sum_{i \in \mathcal{E}(\mathbf{n})^c} \mu_i.$$

Gelenbe [29] shows that the invariant measure of a network with triggered customer movement, subject to the solution of some traffic equations (given as equations (4.22) and (4.23)), is

$$m(\mathbf{n}) = \prod_{i=1}^{N} z_i^{n_i}, \ \text{where } z_i = \frac{\lambda^+_i}{\mu_i + \lambda^s_i}, \tag{4.20}$$

so that the reverse time transition rates required for equations (4.3) to hold are

$$\begin{aligned}
q'(\mathbf{n}, \mathbf{n} + \mathbf{e}_i + \mathbf{e}_j - \mathbf{e}_k) &= \mu_i p^s_{ij} q_{jk} \frac{z_i z_j}{z_k}, \quad i, j \in \mathcal{L},\ k \in \mathcal{E}(\mathbf{n})^c, \\
q'(\mathbf{n}, \mathbf{n} + \mathbf{e}_i - \mathbf{e}_j) &= (\mu_i p^+_{ij} + \nu^s_i q_{ij}) \frac{z_i}{z_j}, \quad i \in \mathcal{L},\ j \in \mathcal{E}(\mathbf{n})^c, \\
q'(\mathbf{n}, \mathbf{n} + \mathbf{e}_i) &= \left(\mu_i p^+_{i0} + \mu_i \sum_{j \in \mathcal{E}(\mathbf{n})} p^s_{ij} \right) z_i, \ i \in \mathcal{L},
\end{aligned}$$

$$\begin{aligned} q'(\mathbf{n},\mathbf{n}-\mathbf{e}_i) &= \frac{\nu_i^+}{z_i}, \quad i \in \mathcal{E}(\mathbf{n})^c, \\ q'(\mathbf{n},\mathbf{n}) &= \sum_{i\in\mathcal{E}(\mathbf{n})} \nu_i^s. \end{aligned} \tag{4.21}$$

Note that $q(\mathbf{n},\mathbf{n}) = q'(\mathbf{n},\mathbf{n})$, $\forall \mathbf{n} \in S$. We now show that $q(\mathbf{n}) = q'(\mathbf{n})$ if and only if the parameters $\{\lambda_i^+, \lambda_i^s\}$ of equation (4.20) satisfy

$$\lambda_i^+ = \sum_{j=1}^N z_j \left(\mu_j p_{ji}^+ + \lambda_j^s q_{ji}\right) + \nu_i^+, \ i \in \mathcal{L}, \tag{4.22}$$

and

$$\lambda_i^s = \sum_{j=1}^N z_j \mu_j p_{ji}^s + \nu_i^s, \ i \in \mathcal{L}. \tag{4.23}$$

The traffic equations (4.22) and (4.23) are the same as those of Gelenbe [29]. Equivalent traffic equations are given by equation (3.8) setting $d_{ij}^+ = p_{ij}^+$, $d_{ij}^s(0) = p_{ij}^s$ and $d_{ij}^s(k) = 0$ if $k \neq 0$.

$$\begin{aligned} & q'(\mathbf{n}) - q'(\mathbf{n},\mathbf{n}) \\ &= \sum_{i\in\mathcal{E}(\mathbf{n})^c} \frac{\nu_i^+}{z_i} + \sum_{i=1}^N z_i \left(\mu_i p_{i0}^+ + \mu_i \sum_{j\in\mathcal{E}(\mathbf{n})} p_{ij}^s\right) \\ &\quad + \sum_{i=1}^N \sum_{j\in\mathcal{E}(\mathbf{n})^c} \frac{z_i}{z_j}(\mu_i p_{ij}^+ + \nu_i^s q_{ij}) + \sum_{i=1}^N \sum_{j=1}^N \sum_{k\in\mathcal{E}(\mathbf{n})^c} \mu_i p_{ij}^s q_{jk} \frac{z_i z_j}{z_k}, \\ &= \sum_{i\in\mathcal{E}(\mathbf{n})^c} \frac{\nu_i^+}{z_i} + \sum_{i=1}^N z_i \mu_i p_{i0}^+ + \sum_{j\in\mathcal{E}(\mathbf{n})^c} \frac{1}{z_j} \left(\lambda_j^+ - \nu_j^+ - \sum_{k=1}^N z_k \lambda_k^s q_{kj}\right) \\ &\quad + \sum_{i=1}^N \sum_{j\in\mathcal{E}(\mathbf{n})^c} \frac{z_i}{z_j} \nu_i^s q_{ij} + \sum_{k\in\mathcal{E}(\mathbf{n})^c} \frac{1}{z_k} \sum_{j=1}^N z_j q_{jk} (\lambda_j^s - \nu_j^s) + \sum_{i=1}^N z_i \mu_i \sum_{j\in\mathcal{E}(\mathbf{n})} p_{ij}^s, \\ &= \sum_{i\in\mathcal{E}(\mathbf{n})^c} (\mu_i + \lambda_i^s) + \sum_{i=1}^N \left(z_i \mu_i p_{i0}^+ + z_i \mu_i \sum_{j\in\mathcal{E}(\mathbf{n})} p_{ij}^s\right), \\ &= \sum_{i\in\mathcal{E}(\mathbf{n})^c} (\mu_i + \nu_i^s) + \sum_{i=1}^N \left(z_i \mu_i p_{i0}^+ + \sum_{j=1}^N z_j \mu_j p_{ji}^s\right), \\ &= \sum_{i\in\mathcal{E}(\mathbf{n})^c} (\mu_i + \nu_i^s) + \sum_{i=1}^N z_i \mu_i \left(1 - \sum_{j=1}^N p_{ij}^+\right), \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i\in\mathcal{E}(\mathbf{n})^c} (\mu_i + \nu_i^s) + \sum_{i=1}^{N} \left(\lambda_i^+ - z_i\lambda_i^s - z_i\mu_i \sum_{j=1}^{N} p_{ij}^+ \right), \\
&= \sum_{i\in\mathcal{E}(\mathbf{n})^c} (\mu_i + \nu_i^s) + \sum_{i=1}^{N} \left(\lambda_i^+ - z_i\lambda_i^s \sum_{j=1}^{N} q_{ij} - z_i\mu_i \sum_{j=1}^{N} p_{ij}^+ \right), \\
&= \sum_{i\in\mathcal{E}(\mathbf{n})^c} (\mu_i + \nu_i^s) + \sum_{i=1}^{N} \left(\lambda_i^+ - \sum_{j=1}^{N} z_j\mu_j p_{ji}^+ - \sum_{j=1}^{N} z_j\lambda_j^s q_{ji} \right), \\
&= \sum_{i\in\mathcal{E}(\mathbf{n})^c} (\mu_i + \nu_i^s) + \sum_{i=1}^{N} \nu_i^+, \\
&= q(\mathbf{n}) - q'(\mathbf{n},\mathbf{n}).
\end{aligned}
$$

Thus we have shown, using a reversibility argument, that the invariant measure of a network with triggered customer movement, as shown in Gelenbe [29], is given by equation (4.20) if and only if equations (4.22) and (4.23) hold.

Remark 4.4.1 *The network with transition rates given by equation (4.21) has a product form invariant measure, given by equation (4.20), as a direct consequence of Lemma 4.3, even though some of the transitions involve simultaneous arrivals to two nodes.*

A trigger instigates an immediate departure from the node at which it arrives, if the node is in a state from which a departure is possible, and subsequent transfer of the triggered customer to another node. Exogenous trigger arrivals (type s "arrivals") may be considered to represent external events that cause an internal transfer of customers, and the routeing of a trigger within the network does not result in any interaction with the outside of the network.

Theorem 4.4.1 *The network of nodes, formed using triggered linking of quasi-reversible nodes, is quasi-reversible with respect to the exogenous arrivals and departures of the entities labelled by* $\mathcal{T} = \{+1, s\}$.

Proof: A type $+1$ departure from the system occurs when a positive customer is routed to the outside of the network following a service completion. Customers which are lost from the network due to the effects of exogenously generated triggers do not constitute type +1 "departures". When discussing quasi-reversibility of the network as a whole we are only interested in exogenous "arrivals" and "departures". Therefore we define any transition which only involves routeing within the network, for example, the internal transfer of a trigger, to be a type $t \notin \mathcal{T}$ event. Set the probabilities $f^a, f^d \in \mathcal{F}$ to be:

$$\begin{aligned} f^a(+1;\mathbf{n},\mathbf{n}+\mathbf{e}_i) &= 1,\ i \in \mathcal{L}, \\ f^d(+1;\mathbf{n}+\mathbf{e}_i,\mathbf{n}) &= \frac{\mu_i p_{i0}^+}{\mu_i p_{i0}^+ + \mu_i \sum_{j\in\mathcal{E}(\mathbf{n})} p_{ij}^s},\ i \in \mathcal{L}, \\ f^a(s;\mathbf{n},\mathbf{n}-\mathbf{e}_i+\mathbf{e}_j) &= \frac{\nu_i^s q_{ij}}{\mu_i p_{ij}^+ + \nu_i^s q_{ij}},\ i \in \mathcal{E}(\mathbf{n})^c,\ j \in \mathcal{L} \\ f^a(s;\mathbf{n},\mathbf{n}) &= 1, \end{aligned} \tag{4.24}$$

with $f^a(t;\mathbf{n},\mathbf{n}')$ and $f^d(t;\mathbf{n},\mathbf{n}')$ zero otherwise. Then from equations (4.4) and (4.5)

$$\alpha(+1,\mathbf{n}) = \sum_{i=1}^{N} \nu_i^+ = \alpha(+1),$$

$$\alpha(s,\mathbf{n}) = \sum_{i=1}^{N} \nu_i^s = \alpha(s),$$

and

$$\beta(+1,\mathbf{n}) = \sum_{i=1}^{N} \mu_i p_{i0}^+ z_i = \beta(+1),$$

so that the arrival and departure streams of type $t \in \mathcal{T}$ entities are state independent. Note that the rates $\alpha(+1)$, $\alpha(s)$ and $\beta(+1)$ are sums over the nodes of the terms ν_i^+, ν_i^s and $\mu_i p_{i0}^+ z_i$ respectively, which are also state independent, so that each node within the network is individually quasi-reversible with respect to exogenous arrivals of type $t \in \mathcal{T}$.

□

In an analogous fashion to the individual quasi-reversible nodes of Section 4.3, $\alpha_i(+1) = \lambda_i^+$, $\alpha_i(s) = \lambda_i^s$ and $\beta_i(+1) = \mu_i z_i$, where λ_i^+ and λ_i^s are the total arrival rates of positive customers and triggers to node i respectively, so that the traffic equations (4.22) and (4.23) become

$$\alpha_i(+1) = \sum_{j=1}^{N} \left(\beta_j(+1)p_{ji}^{+} + z_j\alpha_j(t)q_{ji}\right) + \nu_i^{+}, \ i \in \mathcal{L}, \tag{4.25}$$

and

$$\alpha_i(s) = \left\{\sum_{j=1}^{N} \beta_j(+1)p_{ji}^{s} + \nu_i^{s}\right\}, \ i \in \mathcal{L}. \tag{4.26}$$

Substitute equation (4.26) into equation (4.25) to obtain

$$\alpha_i(+1) = \sum_{j=1}^{N} \left(\beta_j(+1)p_{ji}^{+} + z_j \left\{\sum_{k=1}^{N} \beta_k(+1)p_{kj}^{s} + \nu_j^{s}\right\} q_{ji}\right) + \nu_i^{+}, \ i \in \mathcal{L}. \tag{4.27}$$

The question remains as to how equations (4.26) and (4.27) can be considered to represent the traffic equations necessary for a legitimate linking of nodes, which was our original aim. Equations (4.26) are of the form given by equation (4.9), but equations (4.27) include extra terms, which indicate that a type +1 "departure" at node k (represented by $\beta_k(+1)$) can trigger a decrease in the number of customers at node j (represented by p_{kj}^{s}) and cause a subsequent type +1 "arrival" to node i (represented by q_{ji}), and that exogenous type s "arrivals" to node j (represented by ν_j^{s}) can trigger type +1 "arrivals" to node i. The z_j term places conditions on the effectiveness of the triggering, and will be discussed later. Coupling/linking of quasi-reversible nodes has, until now, been restricted to when a type $t \in \mathcal{T}$ "departure" triggers a type $t' \in \mathcal{T}$ "arrival".

The transition rates for our network with triggered linking, analogous to the transition rates given by equations (4.7) and (4.8) are, for $i \in \mathcal{E}(\mathbf{n})^c$,

$$\begin{aligned} & q(\mathbf{n}, \mathbf{n} - \mathbf{e}_i - \mathbf{e}_j + \mathbf{e}_k) \\ &= q_i(n_i, n_i - 1) f_i^d(+1; n_i, n_i - 1) p_{ij}^{s} I\left\{j \in \mathcal{E}(\mathbf{n} - \mathbf{e}_i)^c\right\} \frac{q_j(n_j, n_j - 1) f_j^d(+1; n_j, n_j - 1)}{\mu_j} \\ & \quad \times q_{jk} \frac{q_k(n_k, n_k + 1) f_k^a(+1; n_k, n_k + 1)}{\alpha_k(+1)}, \ k \in \mathcal{L}, \end{aligned} \tag{4.28}$$

$$\begin{aligned} & q(\mathbf{n}, \mathbf{n} - \mathbf{e}_i + \mathbf{e}_j) \\ &= \nu_i^{s} I\left\{i \in \mathcal{E}(\mathbf{n})^c\right\} \frac{q_i(n_i, n_i - 1) f_i^d(+1; n_i, n_i - 1)}{\mu_i} q_{ij} \frac{q_j(n_j, n_j + 1) f_j^a(+1; n_j, n_j + 1)}{\alpha_j(+1)} \end{aligned}$$

$$+q_i(n_i, n_i-1)f_i^d(+1; n_i, n_i-1)p_{ij}^+ \frac{q_j(n_j, n_j+1)f_j^a(+1; n_j, n_j+1)}{\alpha_j(+1)}, \; j \in \mathcal{L}, \quad (4.29)$$

$q(\mathbf{n}, \mathbf{n}-\mathbf{e}_i)$

$$\begin{aligned} = \; & q_i(n_i, n_i-1)f_i^d(+1; n_i, n_i-1)p_{i0}^+ \\ & + q_i(n_i, n_i-1)f_i^d(+1; n_i, n_i-1)\sum_{j=1}^{N} p_{ij}^s I\{j \in \mathcal{E}(\mathbf{n})\}, \end{aligned} \quad (4.30)$$

and

$$q(\mathbf{n}, \mathbf{n}+\mathbf{e}_i) = \nu_i^+ \frac{q_i(n_i, n_i+1)f_i^a(+1; n_i, n_i+1)}{\alpha_i(+1)}, \; i \in \mathcal{L}, \quad (4.31)$$

$$q(\mathbf{n}, \mathbf{n}) = \sum_{j=1}^{N} \nu_j^s I\{j \in \mathcal{E}(\mathbf{n})\}. \quad (4.32)$$

Equations (4.28) to (4.32) may be seen to be equivalent to equations (4.19) by substituting for $q_i(\cdot,\cdot)$, $f_i^a(\cdot;\cdot,\cdot)$ and $f_i^d(\cdot;\cdot,\cdot)$ from equations (4.18). Let us focus attention on equation (4.28), which has a μ_j and an $\alpha_k(+1)$ term in the denominator, representing the rate of type +1 "departures" from node j and the rate of type +1 "arrivals" to node k respectively. The presence of the indicator function $I\{j \in \mathcal{E}(\mathbf{n}-\mathbf{e}_i)^c\}$ ensures that the rate $q(\mathbf{n}, \mathbf{n}-\mathbf{e}_i-\mathbf{e}_j+\mathbf{e}_k)$ is non-zero only for nodes j from which a type +1 departure can occur when the process is in interim state $\mathbf{n}-\mathbf{e}_i$, that is those nodes which are not empty when the network is in state $\mathbf{n}-\mathbf{e}_i$. Given this conditioning, the rate of type +1 "departures" from node j is μ_j. Note that, under the equilibrium queue length distribution resulting from the invariant measure given by equation (4.20),

$$E\left[I\{j \in \mathcal{E}(\mathbf{n})^c\}\right] = z_j,$$

is the probability that node j is not empty, so that $\beta_j(+1) = z_j\mu_j$ is the *expected* rate of type +1 "departures" from node j under this distribution. This explains why the term in the denominator of equation (4.28) is μ_j rather than $\beta_j(+1)$, which would be analogous to the presence of the $\alpha_k(+1)$ term, and also explains the presence of the z_j term in the traffic equations (4.27). The equations (4.28) to (4.32), governing the triggered linking of nodes, may now be interpreted as follows:

- Equation (4.28) represents transitions in which a type +1 departure from node

i triggers a type +1 "departure" from node $j \in \mathcal{E}(\mathbf{n} - \mathbf{e}_i)^c$ and subsequent type +1 "arrival" at node k.

- Equation (4.29) represents transitions in which a type +1 "departure" from node i triggers a type +1 "arrival" to node j.
- Equation (4.30) represents transitions in which a type +1 "departure" from node i leaves the network, and also transitions in which a type +1 "departure" from node i fails to trigger a type +1 "departure" from node $j \in \mathcal{E}(\mathbf{n})$.
- Equation (4.31) represents transitions in which exogenous type +1 "arrivals" occur at node i.
- Equation (4.32) represents transitions in which exogenous type s "arrivals" fail to have any effect on the network, because they are unable to trigger a type +1 departure from node $j \in \mathcal{E}(\mathbf{n})$.

The mysterious z_j term in equation (4.27) is the expected probability that node j is not empty, and appears to be necessary for equations (4.28) to (4.32) to represent a legitimate linking of nodes.

We have devised a linking in which the queue lengths may change simultaneously at three nodes, and obtained a quasi-reversible network. To our knowledge this has not been achieved before.

Chapter 5

Forced Batch Movement

Each network described thus far restricts positive customers to be routed through the network singly. This requirement has been relaxed by a number of authors who analyse discrete and continuous time queueing networks in which customers arrive and are served in batches. For example, Walrand [72] considers a discrete time open network of queues, with specific forms of service and arrival probabilities, and in which customers are routed independently of the batches in which they find themselves. Henderson, Pearce, Taylor and van Dijk [35] make the same independent routeing assumption, but generalise the service rate/probability form for a closed queueing network.

The assumption of independent routeing restricts the applicability of a model severely. This problem is addressed in Henderson and Taylor [36] and [38], which discuss both open and closed queueing networks with state dependent batch arrival and service probabilities and correlated batch routeing, so that the routeing of customers is dependent on the composition of the released batch of customers. Henderson and Taylor [36] use reverse time arguments, similar to Lemma 4.1.1, to show that their networks have a product form equilibrium distribution, and note that their results are valid for both discrete and continuous time networks.

Boucherie and van Dijk [4] and [5] also describe batch movement queueing networks. In particular, [5] generalises the networks of [36] and [38] to incorporate some degree of state dependence into the batch routeing probabilities.

In parallel, and sometimes directly connected to the results on batch movement queueing networks, product form equilibrium distributions have been found for classes of stochastic Petri nets (SPNs). Henderson, Lucic and Taylor [34] and Henderson and Taylor [37] apply the results of [36] and [38] to the modelling framework of SPNs to derive a class of product form SPNs. Other authors discuss classes of product form SPNs without relating their results to batch movement queueing networks. See, for example, Donatelli and Sereno [21], Frosch and Natarajan [24], and Li and Georganas [56].

The above concepts are merged with the idea of signalling in this chapter by considering a batch movement queueing network (or SPN) within which the triggered release of batches, by batches in transit, is allowed. It is shown that a product form equilibrium distribution still holds.

In Section 5.1 we define the class of queueing networks under consideration in this chapter. Section 5.2 contains the main theorem and proof that this class of networks has a product form equilibrium distribution and Section 5.3 applies the results to stochastic Petri nets (SPNs).

Algorithms which extend those of Buzen [9] and Reiser and Lavenberg [63], for calculating performance measures and normalising constants of product form queueing networks and SPNs, have been derived for the systems described in [34], [36] and [37] by Coleman, Henderson and Taylor in [17] and [18]. The class of SPNs derived in this chapter may be analysed using these algorithms.

We use vector tests of comparison in this chapter. Unless otherwise stated, the tests are done component-wise, for example, $(\mathbf{a} \leq \mathbf{a}')$ is true if and only if each component of $\mathbf{a}$ is less than or equal to the corresponding component of $\mathbf{a}'$.

5.1 System Description

Consider a network consisting of N nodes indexed by the set $\{1, 2, \ldots, N\}$. Associated with each node in the network is a queue of customers. The state space, S, of the network comprises vectors $\mathbf{n} = (n_1, n_2, \ldots, n_N)$ where $n_i \in Z$ is the queue length at node i. Note that we allow for the possibility of negative queue lengths. We also consider there to be an external source of customers (labelled node 0) representing the outside of the network. There is no queue of customers associated with node 0. Customers enter the network by being emitted from node 0 and leave by being absorbed by node 0.

The network behaves according to the following mechanism. When the network is in state $\mathbf{n}$, a non-negative vector $\mathbf{a} = (a_1, a_2, \ldots, a_N)$ of customers is released from the network with a probability (in discrete time) or rate (in continuous time) of $q(\mathbf{n}, \mathbf{a})$. Here a_i is the number of customers released from node i. Denote $\mathcal{A}$ to be the set of all possible vectors which, at some stage, are released from, routed through, or deposited in the network, and note that the null vector $\mathbf{0} = (0, 0, \ldots, 0)$ is allowed to be in $\mathcal{A}$. The release of the null vector signals no change other than the arrival of customers from outside the network, so that it is often the case that $\mathbf{0} \in \mathcal{A}$ whenever an open network is being modelled. This is not of course the only means by which customers can arrive to the network. Any time that the vector $\mathbf{a}$ is released and transformed into $\mathbf{a}'$, with $\sum_{i=1}^{N} a_i' > \sum_{i=1}^{N} a_i$, before being deposited in the network, arrivals will have occurred.

We assume that $q(\mathbf{n}, \mathbf{a})$ has the form

$$q(\mathbf{n}, \mathbf{a}) = \frac{\psi(\mathbf{n} - \mathbf{a})\xi(\mathbf{a})}{\phi(\mathbf{n})}, \tag{5.1}$$

where $\phi(\cdot)$ and $\psi(\cdot)$ satisfy

$$\phi(\mathbf{n}) > 0 \text{ and } \psi(\mathbf{n}) \geq \psi(\mathbf{n} - \mathbf{a}) \;\forall\, \mathbf{n} \in S,\ \mathbf{a} \in \mathcal{A}, \tag{5.2}$$

but otherwise $\phi(\cdot)$, $\psi(\cdot)$ and $\xi(\cdot)$ are arbitrary, but given, non-negative functions defined on Z^N. The choice $\psi(\mathbf{n}-\mathbf{a}) = 0$ whenever $\mathbf{n}-\mathbf{a}$ has negative components ensures that the vector $\mathbf{a}$ cannot be released if doing so would cause any queue lengths to become negative. However this state space restriction does not have to be enforced to retain product form invariant measures, and in some applications it may be suitable for queue lengths to be negative (see [41]).

Henderson [41] and Henderson, Northcote and Taylor [40] also describe how multiple classes of customers may be incorporated into the network by an appropriate choice of the functions $\psi(\cdot)$ and $\phi(\cdot)$.

When $\mathbf{a}$ is released from state $\mathbf{n}$ it will, with probability

$$p(\mathbf{a}, \mathbf{a}', \mathbf{a}''), \text{ for all } \mathbf{a}', \mathbf{a}'' \in \mathcal{A},$$

attempt to force the release of $\mathbf{a}'$ and subsequent deposit of $\mathbf{a}''$. The forcing will be successful with probability

$$\frac{\psi(\mathbf{n}-\mathbf{a}-\mathbf{a}')}{\psi(\mathbf{n}-\mathbf{a})}, \tag{5.3}$$

in which case the new state is $\mathbf{n}-\mathbf{a}-\mathbf{a}'+\mathbf{a}''$. Otherwise the null vector will be deposited and the new state will be $\mathbf{n}-\mathbf{a}$. Equation (5.2) ensures that the acceptance probabilities are well defined.

Batch movement queueing networks without forced batch releases are a special case of these networks, and may be modelled by setting $p(\mathbf{a}, \mathbf{a}', \mathbf{a}'') = 0, \ \forall \mathbf{a}' \neq \mathbf{0}$, so that batches of customers are released, transferred and deposited, without attempting to force the release of any additional customers. The straightforward transfer of batches occurs in networks with forced batch movement when $\mathbf{a}' = \mathbf{0}$, so that the vector $\mathbf{a}$ is released, routed through the network, and subsequently deposited as the vector $\mathbf{a}''$. Notice that in such cases the probability of acceptance given by equation (5.3) is 1, ensuring that $\mathbf{a}''$ will be deposited. For probabilistic routeing there may be a number of vectors $\mathbf{a}''$ satisfying $p(\mathbf{a}, \mathbf{a}', \mathbf{a}'') > 0$.

Without loss of generality we do not allow $p(\mathbf{a}, \mathbf{a}', \mathbf{a}'') > 0$ if $\mathbf{a}+\mathbf{a}' = \mathbf{a}''$, as in such instances there is no net change of state. In particular we set

$$p(\mathbf{a},\mathbf{0},\mathbf{a}) = p(\mathbf{0},\mathbf{a},\mathbf{a}) = 0, \text{ for } \mathbf{a} \in \mathcal{A}, \tag{5.4}$$

and require

$$\sum_{\mathbf{a}'\in\mathcal{A}} \sum_{\mathbf{a}''\in\mathcal{A}} p(\mathbf{a},\mathbf{a}',\mathbf{a}'') = 1, \text{ for } \mathbf{a} \in \mathcal{A}. \tag{5.5}$$

Notice that, when the network is in state $\mathbf{n}$, $q(\mathbf{n},\mathbf{0})p(\mathbf{0},\mathbf{0},\mathbf{a})$ is the rate of arrival of batch $\mathbf{a}$ to the network, and $q(\mathbf{n},\mathbf{0})p(\mathbf{a},\mathbf{0},\mathbf{0})$ is the rate at which batch $\mathbf{a}$ is released and leaves the network due to service completion.

5.2 The Equilibrium Distribution

Theorem 5.2.1 *If there exists a solution* $\{f(\mathbf{a}) > 0,\ \mathbf{a} \in \mathcal{A}\}$*, and* $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ *to the equations*

$$\xi(\mathbf{a})f(\mathbf{a}) = \sum_{\mathbf{a}'\in\mathcal{A}} \sum_{\mathbf{a}''\in\mathcal{A}} \xi(\mathbf{a}')f(\mathbf{a}')\Big[f(\mathbf{a}'')p(\mathbf{a}',\mathbf{a}'',\mathbf{a}) - f(\mathbf{a})p(\mathbf{a}',\mathbf{a},\mathbf{a}'')\Big],\ \mathbf{a} \in \mathcal{A}\backslash\{\mathbf{0}\}, \tag{5.6}$$

and

$$\mathbf{a}\cdot\mathbf{x} = log(f(\mathbf{a})),\ \mathbf{a} \in \mathcal{A}, \tag{5.7}$$

then the queueing network defined in section 5.1 has an invariant measure

$$\begin{aligned} \pi(\mathbf{n}) &= \phi(\mathbf{n})e^{\mathbf{n}\cdot\mathbf{x}}, \\ &= \phi(\mathbf{n})\prod_{i=1}^{N} e^{n_i x_i}. \end{aligned} \tag{5.8}$$

Note that equation (5.7) specifies $f(\mathbf{0}) = 1$*.*

Proof: We prove that equation (5.8), subject to equations (5.6) and (5.7), satisfies the global balance equations for each reachable state in the Markov chain. The global balance equations for state $\mathbf{n}$ are

$$\pi(\mathbf{n})\left[\sum_{\mathbf{a}\in\mathcal{A}} q(\mathbf{n},\mathbf{a}) - q(\mathbf{n},\mathbf{0})\sum_{\mathbf{a}\neq\mathbf{0}}\sum_{\mathbf{a}''\in\mathcal{A}} p(\mathbf{0},\mathbf{a},\mathbf{a}'')\left(1-\frac{\psi(\mathbf{n}-\mathbf{a})}{\psi(\mathbf{n})}\right)\right]$$

$$= \sum_{\mathbf{a}\in\mathcal{A}}\sum_{\mathbf{a}'\in\mathcal{A}}\sum_{\mathbf{a}''\in\mathcal{A}} \pi(\mathbf{n}-\mathbf{a}+\mathbf{a}'+\mathbf{a}'')q(\mathbf{n}-\mathbf{a}+\mathbf{a}'+\mathbf{a}'',\mathbf{a}')p(\mathbf{a}',\mathbf{a}'',\mathbf{a})\frac{\psi(\mathbf{n}-\mathbf{a})}{\psi(\mathbf{n}-\mathbf{a}+\mathbf{a}'')}$$
$$+\sum_{\mathbf{a}\neq\mathbf{0}}\sum_{\mathbf{a}'\neq\mathbf{0}}\sum_{\mathbf{a}''\in\mathcal{A}} \pi(\mathbf{n}+\mathbf{a}')q(\mathbf{n}+\mathbf{a}',\mathbf{a}')p(\mathbf{a}',\mathbf{a},\mathbf{a}'')\left(1-\frac{\psi(\mathbf{n}-\mathbf{a})}{\psi(\mathbf{n})}\right). \qquad (5.9)$$

The LHS of equation (5.9) represents the total flux out of state $\mathbf{n}$. The negative term represents transitions in which $\mathbf{0}$ is emitted from state $\mathbf{n}$ and the resultant forcing attempt is not successful, so that the state of the network does not change. The RHS of equation (5.9) represents the total flux into state $\mathbf{n}$. The first term includes all transitions in which the release of a vector successfully forces the release of an additional vector, and the second term includes all transitions in which the forced release is unsuccessful. Substituting from equations (5.1) and (5.8), dividing throughout by $e^{\mathbf{n}\cdot\mathbf{x}}$ and noting, from equation (5.7), that

$$\frac{e^{(\mathbf{n}+\mathbf{a})\cdot\mathbf{x}}}{e^{\mathbf{n}\cdot\mathbf{x}}} = e^{\mathbf{a}\cdot\mathbf{x}} = f(\mathbf{a}),$$

we obtain

$$\sum_{\mathbf{a}\in\mathcal{A}}\psi(\mathbf{n}-\mathbf{a})\xi(\mathbf{a}) - \psi(\mathbf{n})\xi(\mathbf{0})\sum_{\mathbf{a}\neq\mathbf{0}}\sum_{\mathbf{a}''\in\mathcal{A}} p(\mathbf{0},\mathbf{a},\mathbf{a}'')\left(1-\frac{\psi(\mathbf{n}-\mathbf{a})}{\psi(\mathbf{n})}\right)$$
$$= \sum_{\mathbf{a}\in\mathcal{A}}\sum_{\mathbf{a}'\in\mathcal{A}}\sum_{\mathbf{a}''\in\mathcal{A}} \frac{1}{f(\mathbf{a})}f(\mathbf{a}')f(\mathbf{a}'')\psi(\mathbf{n}-\mathbf{a}+\mathbf{a}'')\xi(\mathbf{a}')p(\mathbf{a}',\mathbf{a}'',\mathbf{a})\frac{\psi(\mathbf{n}-\mathbf{a})}{\psi(\mathbf{n}-\mathbf{a}+\mathbf{a}'')}$$
$$+\sum_{\mathbf{a}\neq\mathbf{0}}\sum_{\mathbf{a}'\neq\mathbf{0}}\sum_{\mathbf{a}''\in\mathcal{A}} f(\mathbf{a}')\psi(\mathbf{n})\xi(\mathbf{a}')p(\mathbf{a}',\mathbf{a},\mathbf{a}'')\left(1-\frac{\psi(\mathbf{n}-\mathbf{a})}{\psi(\mathbf{n})}\right).$$

We rearrange this last equation to get all of the $\psi(\mathbf{n}-\mathbf{a})$ terms on the LHS and all of the $\psi(\mathbf{n})$ terms on the RHS, remembering that $f(\mathbf{0})=1$, to obtain

$$\sum_{\mathbf{a}\neq\mathbf{0}}\psi(\mathbf{n}-\mathbf{a})\left[\xi(\mathbf{a}) - \sum_{\mathbf{a}'\in\mathcal{A}}\sum_{\mathbf{a}''\in\mathcal{A}} f(\mathbf{a}')\xi(\mathbf{a}')\left\{\frac{1}{f(\mathbf{a})}f(\mathbf{a}'')p(\mathbf{a}',\mathbf{a}'',\mathbf{a}) - p(\mathbf{a}',\mathbf{a},\mathbf{a}'')\right\}\right]$$
$$= \psi(\mathbf{n})\left[-\xi(\mathbf{0}) + \xi(\mathbf{0})\sum_{\mathbf{a}\neq\mathbf{0}}\sum_{\mathbf{a}''\in\mathcal{A}} p(\mathbf{0},\mathbf{a},\mathbf{a}'') + \sum_{\mathbf{a}'\in\mathcal{A}}\sum_{\mathbf{a}''\in\mathcal{A}} f(\mathbf{a}')f(\mathbf{a}'')\xi(\mathbf{a}')p(\mathbf{a}',\mathbf{a}'',\mathbf{0})\right.$$
$$\left.+\sum_{\mathbf{a}\neq\mathbf{0}}\sum_{\mathbf{a}'\neq\mathbf{0}}\sum_{\mathbf{a}''\in\mathcal{A}} f(\mathbf{a}')\xi(\mathbf{a}')p(\mathbf{a}',\mathbf{a},\mathbf{a}'')\right].$$

Collecting terms and simplifying we obtain

$$\sum_{\mathbf{a}\neq\mathbf{0}}\psi(\mathbf{n}-\mathbf{a})\left[\xi(\mathbf{a}) - \sum_{\mathbf{a}'\in\mathcal{A}}\sum_{\mathbf{a}''\in\mathcal{A}} f(\mathbf{a}')\xi(\mathbf{a}')\left\{\frac{1}{f(\mathbf{a})}f(\mathbf{a}'')p(\mathbf{a}',\mathbf{a}'',\mathbf{a}) - p(\mathbf{a}',\mathbf{a},\mathbf{a}'')\right\}\right]$$

$$= \psi(\mathbf{n}) \left[-\xi(\mathbf{0}) + \sum_{\mathbf{a}' \in \mathcal{A}} \sum_{\mathbf{a}'' \in \mathcal{A}} f(\mathbf{a}')\xi(\mathbf{a}') \left\{ f(\mathbf{a}'')p(\mathbf{a}', \mathbf{a}'', \mathbf{0}) + \sum_{\mathbf{a} \neq \mathbf{0}} p(\mathbf{a}', \mathbf{a}, \mathbf{a}'') \right\} \right]. \tag{5.10}$$

Using equation (5.6) the LHS of equation (5.10) is zero leaving us to show that

$$\xi(\mathbf{0}) = \sum_{\mathbf{a}' \in \mathcal{A}} \sum_{\mathbf{a}'' \in \mathcal{A}} f(\mathbf{a}')\xi(\mathbf{a}') \left\{ f(\mathbf{a}'')p(\mathbf{a}', \mathbf{a}'', \mathbf{0}) + \sum_{\mathbf{a} \neq \mathbf{0}} p(\mathbf{a}', \mathbf{a}, \mathbf{a}'') \right\}. \tag{5.11}$$

We do so by summing equation (5.6) over all $\mathbf{a} \neq \mathbf{0}$ as follows

$$\sum_{\mathbf{a} \neq \mathbf{0}} \xi(\mathbf{a}) f(\mathbf{a}) = \sum_{\mathbf{a} \neq \mathbf{0}} \sum_{\mathbf{a}' \in \mathcal{A}} \sum_{\mathbf{a}'' \in \mathcal{A}} \xi(\mathbf{a}') f(\mathbf{a}') \Big[f(\mathbf{a}'')p(\mathbf{a}', \mathbf{a}'', \mathbf{a}) - f(\mathbf{a})p(\mathbf{a}', \mathbf{a}, \mathbf{a}'') \Big].$$

Using $f(\mathbf{0}) = 1$,

$$\sum_{\mathbf{a} \in \mathcal{A}} \xi(\mathbf{a}) f(\mathbf{a}) = \xi(\mathbf{0}) + \sum_{\mathbf{a} \neq \mathbf{0}} \sum_{\mathbf{a}' \in \mathcal{A}} \sum_{\mathbf{a}'' \in \mathcal{A}} \xi(\mathbf{a}') f(\mathbf{a}') \Big[f(\mathbf{a}'')p(\mathbf{a}', \mathbf{a}'', \mathbf{a}) - f(\mathbf{a})p(\mathbf{a}', \mathbf{a}, \mathbf{a}'') \Big].$$

Therefore, changing the dummy index in the LHS sum to $\mathbf{a}'$, we obtain,

$$\begin{aligned} \xi(\mathbf{0}) &= \sum_{\mathbf{a}' \in \mathcal{A}} \xi(\mathbf{a}') f(\mathbf{a}') \left\{ 1 - \sum_{\mathbf{a} \neq \mathbf{0}} \sum_{\mathbf{a}'' \in \mathcal{A}} f(\mathbf{a}'')p(\mathbf{a}', \mathbf{a}'', \mathbf{a}) + \sum_{\mathbf{a} \neq \mathbf{0}} \sum_{\mathbf{a}'' \in \mathcal{A}} f(\mathbf{a})p(\mathbf{a}', \mathbf{a}, \mathbf{a}'') \right\}, \\ &= \sum_{\mathbf{a}' \in \mathcal{A}} \xi(\mathbf{a}') f(\mathbf{a}') \left\{ 1 - \sum_{\mathbf{a} \neq \mathbf{0}} \sum_{\mathbf{a}'' \in \mathcal{A}} f(\mathbf{a}'')p(\mathbf{a}', \mathbf{a}'', \mathbf{a}) + \sum_{\mathbf{a} \in \mathcal{A}} \sum_{\mathbf{a}'' \neq \mathbf{0}} f(\mathbf{a}'')p(\mathbf{a}', \mathbf{a}'', \mathbf{a}) \right\}. \end{aligned}$$

We now include and remove terms in the RHS {} to make both double sums over $\mathbf{a} \in \mathcal{A}$ and $\mathbf{a}'' \in \mathcal{A}$, so that

$$\begin{aligned} \xi(\mathbf{0}) &= \sum_{\mathbf{a}' \in \mathcal{A}} \xi(\mathbf{a}') f(\mathbf{a}') \left\{ 1 - \sum_{\mathbf{a} \in \mathcal{A}} \sum_{\mathbf{a}'' \in \mathcal{A}} f(\mathbf{a}'')p(\mathbf{a}', \mathbf{a}'', \mathbf{a}) + \sum_{\mathbf{a} \in \mathcal{A}} \sum_{\mathbf{a}'' \in \mathcal{A}} f(\mathbf{a}'')p(\mathbf{a}', \mathbf{a}'', \mathbf{a}) \right. \\ &\qquad \left. + \sum_{\mathbf{a}'' \in \mathcal{A}} f(\mathbf{a}'')p(\mathbf{a}', \mathbf{a}'', \mathbf{0}) - \sum_{\mathbf{a} \in \mathcal{A}} f(\mathbf{0})p(\mathbf{a}', \mathbf{0}, \mathbf{a}) \right\}, \\ &= \sum_{\mathbf{a}' \in \mathcal{A}} \xi(\mathbf{a}') f(\mathbf{a}') \left\{ 1 + \sum_{\mathbf{a}'' \in \mathcal{A}} f(\mathbf{a}'')p(\mathbf{a}', \mathbf{a}'', \mathbf{0}) - \sum_{\mathbf{a} \in \mathcal{A}} f(\mathbf{0})p(\mathbf{a}', \mathbf{0}, \mathbf{a}) \right\}, \\ &= \sum_{\mathbf{a}' \in \mathcal{A}} \xi(\mathbf{a}') f(\mathbf{a}') \left\{ \sum_{\mathbf{a}'' \in \mathcal{A}} f(\mathbf{a}'')p(\mathbf{a}', \mathbf{a}'', \mathbf{0}) + \sum_{\mathbf{a} \in \mathcal{A}} \sum_{\mathbf{a}'' \neq \mathbf{0}} p(\mathbf{a}', \mathbf{a}'', \mathbf{a}) \right\}, \\ &= \sum_{\mathbf{a}' \in \mathcal{A}} \sum_{\mathbf{a}'' \in \mathcal{A}} \xi(\mathbf{a}') f(\mathbf{a}') \left\{ f(\mathbf{a}'')p(\mathbf{a}', \mathbf{a}'', \mathbf{0}) + \sum_{\mathbf{a} \neq \mathbf{0}} p(\mathbf{a}', \mathbf{a}, \mathbf{a}'') \right\}, \end{aligned}$$

and thus equation(5.11) and the GBEs are satisfied.

□

Theorem 5.2.2 *There always exists a non-negative solution* $\{f(\mathbf{a}),\ \mathbf{a} \in \mathcal{A}\}$ *to equations (5.6).*

Proof: We can construct a single movement queueing network, with constant firing rates and triggered customer movement, as in Gelenbe [29], for which equations (5.6) are the same traffic equations which must hold for the constructed network to have a geometric product form invariant measure. We construct the network as follows. Let there be one queue in the constructed network for each $\mathbf{a} \in \mathcal{A}$. The service rate of customers at queue $\mathbf{a}$ is $\xi(\mathbf{a})$. Once served at queue $\mathbf{a}$ a customer will trigger the release of a customer from queue $\mathbf{a}'$ and proceed to queue $\mathbf{a}''$ with probability $p(\mathbf{a}, \mathbf{a}', \mathbf{a}'')$. If no customer is present at queue $\mathbf{a}'$ then the customer released from queue $\mathbf{a}$ leaves the network. Queue $\mathbf{0}$ represents the outside of the network. The state of this network is $\mathbf{n} = (n(\mathbf{a}); \mathbf{a} \in \mathcal{A})$, where $n(\mathbf{a})$ is the number of customers in queue $\mathbf{a}$. The invariant measure and traffic equations for this network are, with appropriate changes in notation (for example $p^-(\mathbf{a}, \mathbf{a}')q(\mathbf{a}', \mathbf{a}'')$ becomes $p(\mathbf{a}, \mathbf{a}', \mathbf{a}'')$), the same as equations (1.8) and (1.10). After substituting for $\lambda^-(\cdot)$ in the equation for $\lambda^+(\cdot)$ (to combine the $p^-(\cdot, \cdot)$ and $q(\cdot, \cdot)$ terms), remembering that $f(\mathbf{0}) = 1$, the traffic equations are, for $\mathbf{a} \in \mathcal{A}\backslash\{\mathbf{0}\}$,

$$\begin{aligned}
\lambda^+(\mathbf{a}) &= \sum_{\mathbf{a}' \neq \mathbf{0}} f(\mathbf{a}') \left(\xi(\mathbf{a}')p(\mathbf{a}', \mathbf{0}, \mathbf{a}) + \sum_{\mathbf{a}'' \neq \mathbf{0}} f(\mathbf{a}'')\xi(\mathbf{a}'')p(\mathbf{a}'', \mathbf{a}', \mathbf{a}) + \xi(\mathbf{0})p(\mathbf{0}, \mathbf{a}', \mathbf{a}) \right) \\
&\quad + \xi(\mathbf{0})p(\mathbf{0}, \mathbf{0}, \mathbf{a}), \\
&= \sum_{\mathbf{a}' \in \mathcal{A}} \sum_{\mathbf{a}'' \in \mathcal{A}} \xi(\mathbf{a}')f(\mathbf{a}')f(\mathbf{a}'')p(\mathbf{a}', \mathbf{a}'', \mathbf{a}), \qquad (5.12)
\end{aligned}$$

and

$$\begin{aligned}
\lambda^-(\mathbf{a}) &= \sum_{\mathbf{a}' \neq \mathbf{0}} f(\mathbf{a}') \sum_{\mathbf{a}'' \in \mathcal{A}} \xi(\mathbf{a}')p(\mathbf{a}', \mathbf{a}, \mathbf{a}'') + \xi(\mathbf{0}) \sum_{\mathbf{a}'' \in \mathcal{A}} p(\mathbf{0}, \mathbf{a}, \mathbf{a}''), \\
&= \sum_{\mathbf{a}' \in \mathcal{A}} \sum_{\mathbf{a}'' \in \mathcal{A}} \xi(\mathbf{a}')f(\mathbf{a}')p(\mathbf{a}', \mathbf{a}, \mathbf{a}''), \qquad (5.13)
\end{aligned}$$

where, from equation (1.8),

$$f(\mathbf{a}) = \frac{\lambda^+(\mathbf{a})}{\xi(\mathbf{a}) + \lambda^-(\mathbf{a})}. \tag{5.14}$$

Gelenbe [29] proves that there always exists a non-negative solution $\{f(\mathbf{a}),\ \mathbf{a} \in \mathcal{A}\}$ to equations (5.12), (5.13) and (5.14). Equation (5.14) may be rearranged to give

$$\lambda^+(\mathbf{a}) = f(\mathbf{a})\xi(\mathbf{a}) + f(\mathbf{a})\lambda^-(\mathbf{a}).$$

Thus, from equations (5.12) and (5.13), for $\mathbf{a} \in \mathcal{A}\backslash\{\mathbf{0}\}$,

$$\sum_{\mathbf{a}'\in\mathcal{A}}\sum_{\mathbf{a}''\in\mathcal{A}} \xi(\mathbf{a}')f(\mathbf{a}')f(\mathbf{a}'')p(\mathbf{a}',\mathbf{a}'',\mathbf{a}) = \xi(\mathbf{a})f(\mathbf{a}) + f(\mathbf{a})\sum_{\mathbf{a}'\in\mathcal{A}}\sum_{\mathbf{a}''\in\mathcal{A}} \xi(\mathbf{a}')f(\mathbf{a}')p(\mathbf{a}',\mathbf{a},\mathbf{a}''),$$

and therefore equations (5.6) must have a non-negative solution $\{f(\mathbf{a}),\ \mathbf{a} \in \mathcal{A}\}$.

□

Note that when the Markov chain for the constructed network is positive recurrent, any solution $\{f(\mathbf{a}),\ \mathbf{a} \in \mathcal{A}\}$ to equations (5.6) is strictly positive, and an invariant measure of the constructed network is

$$\pi(\mathbf{n}) = \prod_{\mathbf{a}\in\mathcal{A}} f(\mathbf{a})^{n(\mathbf{a})}.$$

In this case the $\{f(\mathbf{a}),\ \mathbf{a} \in \mathcal{A}\}$ must be unique because they constitute an invariant measure for a positive recurrent Markov chain.

Remark 5.2.1 *Obtaining a solution $\{f(\mathbf{a}),\ \mathbf{a} \in \mathcal{A}\}$ to equations (5.6) does not guarantee an invariant measure for the network. We also require that $f(\mathbf{a}) > 0$ which imposes strong conditions on the structure of the queueing network. For example when no forced releases occur in the queueing network (see [36]) equations (5.6) are the equilibrium equations for a Markov chain and have a positive solution only if the chain is positive recurrent.*

Define the sets of release, forced and deposit vectors to be

$$\begin{aligned}
\mathcal{R} &= \{\mathbf{a} \in \mathcal{A}\backslash\{\mathbf{0}\} : \xi(\mathbf{a}) > 0, \}, \\
\mathcal{F} &= \{\mathbf{a}' \in \mathcal{A}\backslash\{\mathbf{0}\} : \xi(\mathbf{a})p(\mathbf{a},\mathbf{a}',\mathbf{a}'') > 0,\ \text{for some } \mathbf{a},\mathbf{a}'' \in \mathcal{A}\}, \\
\mathcal{D} &= \{\mathbf{a}'' \in \mathcal{A}\backslash\{\mathbf{0}\} : \xi(\mathbf{a})p(\mathbf{a},\mathbf{a}',\mathbf{a}'') > 0,\ \text{for some } \mathbf{a},\mathbf{a}' \in \mathcal{A}\}.
\end{aligned}$$

Theorem 5.2.3 *There exists a solution* $\{f(\mathbf{a}) > 0, \mathbf{a} \in \mathcal{A}\}\backslash\{\mathbf{0}\}$ *to equations (5.6) only if the following conditions are met:*

- *Each release vector must be a deposit vector for some transition which occurs with positive rate, so that* $\mathbf{a} \in \mathcal{R} \Longrightarrow \mathbf{a} \in \mathcal{D}$.
- *Each forced vector must be a deposit vector for some transition which occurs with positive rate, so that* $\mathbf{a} \in \mathcal{F} \Longrightarrow \mathbf{a} \in \mathcal{D}$.
- *Each deposit vector must be either a release vector or a forced vector for some transition which occurs with positive rate, so that* $\mathbf{a} \in \mathcal{D} \Longrightarrow \mathbf{a} \in (\mathcal{R} \bigcup \mathcal{F})$.

These conditions may be summarised to be $\{\mathbf{a} \in \mathcal{D} \Longleftrightarrow \mathbf{a} \in (\mathcal{R} \bigcup \mathcal{F}),\ \forall \mathbf{a} \in \mathcal{A}\backslash\{\mathbf{0}\}\}$. *We call this the release-deposit condition.*

Proof: Assume that $f(\mathbf{a}) > 0,\ \forall \mathbf{a} \in \mathcal{A}$. If $\mathbf{a} \in \mathcal{R}$ and $\mathbf{a} \neq \mathbf{0}$ then the LHS of equation (5.6) is positive, and so the first term on the RHS of the equation must also be positive, which implies that $\mathbf{a} \in \mathcal{D}$. Similarly, if $\mathbf{a} \in \mathcal{F}$ and $\mathbf{a} \neq \mathbf{0}$ then the second term on the RHS is non-zero, requiring the first term to be positive, implying that $\mathbf{a} \in \mathcal{D}$. The reverse argument holds also, so that if $\mathbf{a} \in \mathcal{D}$ and $\mathbf{a} \neq \mathbf{0}$ then, necessarily, $\mathbf{a} \in (\mathcal{R} \bigcup \mathcal{F})$.

$\square$

The case $\{\mathbf{0} \in \mathcal{A},\ \xi(\mathbf{0}) > 0\}$ describes an open network. We do not include $\mathbf{0}$ in the sets $\mathcal{R}$, $\mathcal{F}$ and $\mathcal{D}$ in such cases, as it is a special release vector. Equation (5.11) is the only constraint which must hold for the vector $\mathbf{0}$ to ensure that the invariant measure of the system is given by equation (5.8). Note that equation (5.11) does not imply $\xi(\mathbf{a})p(\mathbf{a}, \mathbf{0}\,\mathbf{a}'') > 0$, for some $\mathbf{a}, \mathbf{a}'' \in \mathcal{A}$, or $\xi(\mathbf{a})p(\mathbf{a}, \mathbf{a}', \mathbf{0}) > 0$, for some $\mathbf{a}, \mathbf{a}' \in \mathcal{A}$, which are the conditions required for $\mathbf{0} \in \mathcal{F}$ and $\mathbf{0} \in \mathcal{D}$ respectively. It does however imply that if $\xi(\mathbf{0}) = 0$, then $p(\mathbf{a}', \mathbf{a}'', \mathbf{0}) = 0$ for all $\mathbf{a}', \mathbf{a}'' \in \mathcal{A}$, and $p(\mathbf{a}', \mathbf{a}, \mathbf{a}'') = 0$

for all $\mathbf{a} \in \mathcal{A}\backslash\{\mathbf{0}\}$, which represents a network in which forced batch releases may not occur.

Remark 5.2.2 *Theorem 5.2.3 shows that the release-deposit condition is necessary, but not sufficient, for there to be a positive solution to equations (5.6).*

For example, Figure 5.1 is the routeing chain for a system which satisfies the release-deposit condition, but in which, for at least one $\mathbf{a} \in \mathcal{A}$, $f(\mathbf{a}) = 0$. The system has no forced batch movement, but is a special case of the model described in Section 5.1. The labelled states of the routeing chain form the set $\mathcal{A}$, and the directed arcs between them indicate which probabilities $p(\mathbf{a}, \mathbf{0}, \mathbf{a}')$ are non zero. For each arc the release vector $\mathbf{a}$ and deposit vector $\mathbf{a}'$ are at the base and arrow end of the arc respectively.

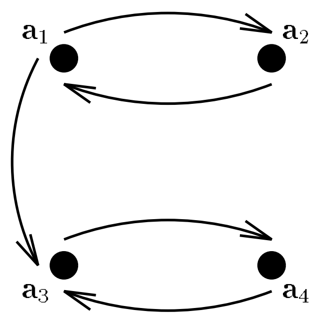

Figure 5.1: An example of a transient routeing chain

Batch movement queueing networks without forced batch movement have transition probabilities $\{p(\mathbf{a}, \mathbf{a}', \mathbf{a}'') > 0$ only if $\mathbf{a}' = \mathbf{0}\}$. In this case equations (5.6) become

$$\xi(\mathbf{a})f(\mathbf{a}) = \sum_{\mathbf{a}' \in \mathcal{A}} \xi(\mathbf{a}')f(\mathbf{a}')p(\mathbf{a}', \mathbf{0}, \mathbf{a}), \;\; \mathbf{a} \in \mathcal{A},$$

which are the traffic equations for the networks of Henderson and Taylor [36]. These traffic equations have a solution $\{f(\mathbf{a}) > 0, \;\; \mathbf{a} \in \mathcal{A}\}$, if and only if the routeing chain, defined by the $p(\mathbf{a}, \mathbf{0}, \mathbf{a}'')$ probabilities, is positive recurrent. It is easy to check the validity of this requirement for a given network. For positive recurrence it must be possible to communicate (along the directed arcs) between every pair of points in the routeing chain diagram. It is obvious that the routeing chain in Figure 5.1 is not positive recurrent. The solution to equations (5.6) for this system are $f(\mathbf{a}_1) = f(\mathbf{a}_2) = 0,\ \xi(\mathbf{a}_3)f(\mathbf{a}_3) = \xi(\mathbf{a}_4)f(\mathbf{a}_4)$.

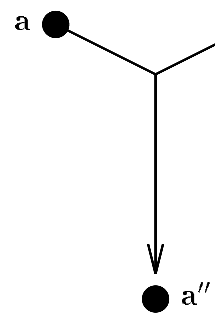

Figure 5.2: An arc that represents forced batch movement

We wish to generalise this for queueing networks with forced batch movement. To do this we introduce a new type of arc, drawn in Figure 5.2, with two source points ($\mathbf{a}$ and $\mathbf{a}'$) and one destination point ($\mathbf{a}''$) to represent the $p(\mathbf{a}, \mathbf{a}', \mathbf{a}'')$ transitions. In the following we conjecture that the resulting "transition graph", (strictly speaking it is not a routeing chain diagram), may be used to examine the positive recurrence of the routeing chain.

Conjecture 5.2.1 *If the "transition graph" of a network, with one point for each* $\mathbf{a} \in \mathcal{A}$ *and one arc (with two sources) for each* $p(\mathbf{a}, \mathbf{a}', \mathbf{a}'') > 0$*, is fully connected so that it is possible to communicate between every pair of points, then equations (5.6) will have a solution* $\{f(\mathbf{a}) > 0\ \forall \mathbf{a} \in \mathcal{A}\}$*.*

Clearly then, if the network satisfies this requirement, it will also satisfy the release-deposit condition of Theorem 5.2.3.

Remark 5.2.3 *Equation (5.7) requires*

$$rank(A) = rank(A|\mathbf{f}), \tag{5.15}$$

where A *is a matrix whose rows are each* $\mathbf{a} \in \mathcal{A}$*,* $\mathbf{f}$ *is a vector with entries* $log(f(\mathbf{a}))$ *in the same order as the rows of* A*, and* $A|\mathbf{f}$ *is the matrix* A *augmented with an additional* *column, namely* $\mathbf{f}$*. It is equivalent to the rank condition of Coleman, Henderson and Taylor [18]. Having found a solution to equations (5.6), a simple test, given by equation (5.15), may be used to determine whether the queueing network under consideration has a product form invariant measure, given by equation (5.8). There are many examples in which equation (5.7) has a solution in* $\mathbf{x}$ *for particular, but not arbitrary, parameter values of the queueing network. In other cases the product form of equations (5.8) is valid for all parameter values as, for example, in Jackson networks of queues [46].*

Remark 5.2.4 *Having determined the existence of a solution* $\{f(\mathbf{a}),\ \mathbf{a} \in \mathcal{A}\}$ *to equations (5.6), the invariant measure of the queueing network in question will lead to an equilibrium distribution if it can be normalised. This requires* $f(\mathbf{a}) < 1$*, for* $\mathbf{a} \in \mathcal{A} \backslash \{\mathbf{0}\}$*. If the equilibrium distribution does exist then* $f(\mathbf{a})$ *may be interpreted as the probability that the network is in a state from which* $\mathbf{a}$ *may be released.*

5.3 Applications

5.3.1 Stochastic Petri Nets with Constant Firing Rates

One important application of batch movement queueing networks is stochastic Petri nets (SPNs). The reader is referred to Peterson [59] for a general introduction to the definition and properties of SPNs. The results here follow from the observation that the marking process of a SPN is essentially a batch movement queueing network.

We adopt an approach, similar to the adaption of [36] and [38] used in [34] and [37], to derive a class of SPNs with a product form equilibrium distribution, in which the firing of a transition can force the immediate firing of a second transition and divert the resultant output bag of tokens.

Consider a SPN, whose underlying Markov process is stationary, with a finite set $\mathcal{P} = \{1, 2, \ldots, N\}$ of places and a finite set $\mathcal{T}$ of transitions. The Markov process representing the SPN has markings $\mathbf{m} = (m_1, m_2, \ldots, m_N) \in Z_+^N$, where m_i is the number of tokens in place i.

Define

$$q(\mathbf{m}, \mathbf{I}(t)) = \psi(\mathbf{m} - \mathbf{I}(t))\xi(t), \tag{5.16}$$

where $\mathbf{I}(t)$ is the input bag of transition t,

$$\psi(\mathbf{m}) = \begin{cases} 1, & \text{if } \mathbf{m} \geq \mathbf{0}, \\ 0, & \text{otherwise,} \end{cases} \tag{5.17}$$

and $\xi(t)$ is the firing rate parameter for transition t.

The SPN behaves according to the following mechanism. When transition t fires it releases the input bag of tokens $\mathbf{I}(t) \in Z_+^N$ from the net, leaving behind a temporary residual marking of $\mathbf{m} - \mathbf{I}(t)$. If, for $t \in \mathcal{T}$, $\mathbf{m} \geq \mathbf{I}(t)$ then transition t has a state independent firing rate of $q(\mathbf{m}, \mathbf{I}(t)) = \xi(t)$. Note that if $\mathbf{m} < \mathbf{I}(t)$ for at least one component of $\mathbf{m}$ then $\psi(\mathbf{m} - \mathbf{I}(t)) = 0 \Longrightarrow q(\mathbf{m}, \mathbf{I}(t)) = 0$ so that transition t cannot fire. Without loss of generality we can assume that there is a one to one correspondence between input bags and transitions, so that no two transitions may have the same input bag. We can model a situation in which two transitions have the same input bag by merging the transitions and using probabilistic routeing (for further discussion see Henderson, Lucic and Taylor [34]). When a transition fires, the released input bag is transformed into an output bag and deposited in the Petri net. We allow each transition to have a number of possible output bags. i.e. we assume that the output bag is chosen according to a probability distribution. This assumption parallels the usual assumption in queueing networks that a served customer is routed to a new queue, or out of the network, according to a probability distribution.

Our new class of product form SPNs have the common feature of "forced transition firing" which we define to happen when a transition, having fired, forces another transition to fire even though the lifetime governing the firing of the second transition has not yet expired. The forced firing of the second transition may deposit a different output bag to that which would have been deposited had the transition fired of its own accord.

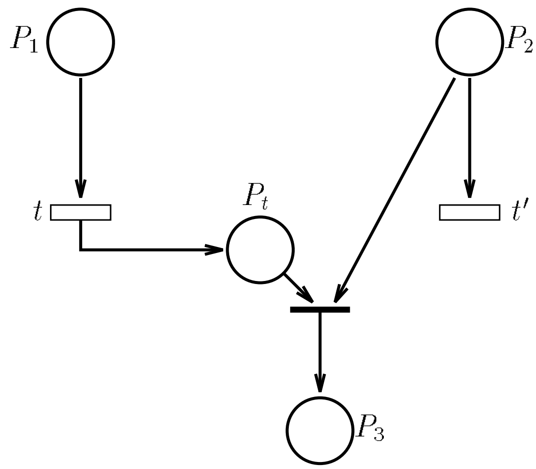

Figure 5.3: Delayed forced transition firing

We apply the forcing concepts of Section 5.1 to these Petri nets to obtain a new class of Petri nets with product form equilibrium distributions. Consider now a Petri net, as described above, in which the release of one input bag, $\mathbf{I}(t)$ say, due to the firing of transition t may force the release of an additional vector of tokens, $\mathbf{I}(t')$ say, and subsequent deposition of a vector of tokens, $\mathbf{O}(t,t')$ say. Noting that if $\mathbf{I}(t')$ is the input bag for transition t', its forced release may be thought of as the forced firing of transition t' by the firing of transition t. Note also that the vector $\mathbf{O}(t,t')$ is not necessarily one of the regular output bags of either transition t or transition t'. Equation (5.17) ensures that the the forced release will not occur unless $\mathbf{I}(t')$ is present in the residual marking ($\mathbf{m}-\mathbf{I}(t) \geq \mathbf{I}(t')$). This forcing concept has been illustrated in Figure 5.3 in which the firing of transition t causes the release of $\mathbf{I}(t)$ and deposit of a token in place P_t. The token deposited in P_t will enable the immediate release of itself and $\mathbf{I}(t')$, if $\mathbf{I}(t')$ is present in the residual marking $\mathbf{m}-\mathbf{I}(t)$, and subsequent deposit of a token in place P_3. We call P_t a transit or temporary place. If $\mathbf{I}(t')$ is successfully released then transition t has, in effect forced transition t' to fire and diverted the resultant output bag. Note that the regular output bags of transitions t and t' are not shown in Figures 5.3 or 5.4.

Figure 5.3 is not yet analogous to the forcing transitions of the batch movement queueing network described in Section 5.1, as tokens deposited in P_t will remain there

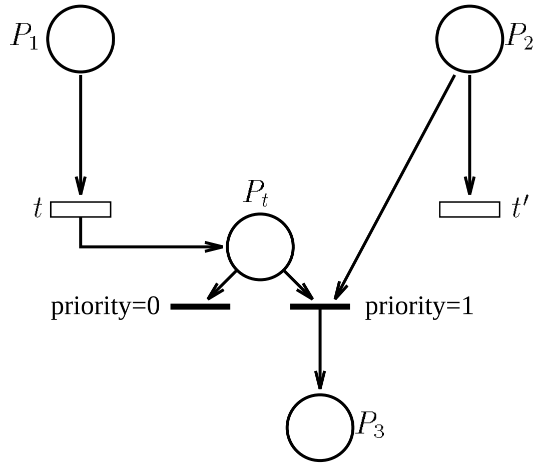

Figure 5.4: Forced transition firing with conflict resolution

until they are successful in forcing the release of $\mathbf{I}(t')$. The forced firing of transition t' by transition t will be delayed if $\mathbf{I}(t')$ is not present in the residual marking. To complete the analogy we require the token, placed in P_t by the firing of transition t, to be lost immediately from the net if $\mathbf{I}(t')$ is not immediately released. This can be implemented in a Petri net by assigning priorities to the conflicting transitions, as in Figure 5.4, in which it is possible for two immediate transitions to become enabled whenever a token is deposited in place P_t. When more than one immediate transition is enabled, the one with the highest priority will fire. In Figure 5.4 the immediate transition with priority 1 will fire if both immediate transitions are enabled, and the one with priority 0 whenever only P_t has a token. Hence the token deposited in place P_t when transition t fires is immediately removed from P_t regardless of whether or not the forcing of transition t' is successful, and that a token is placed in P3 ($\mathbf{O}(t, t')$ is deposited) if and only if transition t is successful in forcing the firing of transition t' (release of $\mathbf{I}(t')$).

We now proceed to formalise the class of Petri nets with forced transition firing. When transition t is enabled in marking $\mathbf{m}$, it fires at rate $\xi(t)$, releasing input bag $\mathbf{I}(t)$. It will, with probability

$$p(\mathbf{I}(t), \mathbf{I}(t'), \mathbf{O}(t, t')) \text{ for } t, t' \in \mathcal{T},$$

attempt to force the release of $\mathbf{I}(t')$ and subsequent deposit of an output bag which is dependent on both t and t', namely $\mathbf{O}(t,t')$. $\mathbf{O}(t,t')$ will be deposited only if the attempted forcing of t' is successful. The forcing will be successful if $\mathbf{m} - \mathbf{I}(t) \geq \mathbf{I}(t')$. The network will be left in state $\mathbf{m} - \mathbf{I}(t) - \mathbf{I}(t') + \mathbf{O}(t,t')$ if the forcing attempt is successful and $\mathbf{m} - \mathbf{I}(t)$ otherwise. We define a null transition t_0, with input bag $\mathbf{I}(t_0) = \mathbf{0}$, and consider the probability $p(\mathbf{I}(t), \mathbf{0}, \mathbf{O}(t,t_0))$ to represent situations when transition t behaves as a normal transition and does not force any other transition to fire. In these circumstances $\mathbf{O}(t,t_0)$ is a regular output bag of t. Without loss of generality we say that $t_0 \in \mathcal{T}$.

We define $\mathcal{A} = \bigcup_{t\in\mathcal{T}} \{\mathbf{I}(t) \bigcup_{t'\in\mathcal{T}} \mathbf{O}(t,t')\}$ to be the finite set of all possible input bags and output bags, so that we may apply the results of Section 5.2 to obtain a class of SPNs with product form invariant measure given by equation (5.8).

Note that the condition given in Theorem 5.2.3 is always satisfied for these SPNs, because we restrict all vectors which are forcibly released to be the input bag of some transition $t \in \mathcal{T}$. It may occur that the vector $\mathbf{I}(t)$ is always forcibly released, in which case transition t never fires. This may be modelled by setting $\xi(t) = 0$ and so, for example, in Figure 5.4 we could set $\xi(t') = 0$ making transition t' redundant.

An Example

Consider the SPN in Figure 5.5 which consists of two timed transitions (T_1 and T_2) with non-trivial input bags, and one (T0) which has $\mathbf{0}$ as its input bag. The firing rates are given on the right of each transition. All arcs in this example have multiplicity one with probabilistic output bags indicated by labelling of the arcs when applicable. As soon as a token arrives at place P_t it is removed because at least one of the immediate transitions is enabled. Thus tokens do not build up in place P_t and so it may be regarded as being a temporary place. When T_0 fires it deposits a single token in P_1 or P_2 with probability α and $1-\alpha$ respectively. When T_1 fires it will, with probability β, behave as a normal transition and deposit a token in P_3. With probability $1-\beta$ the firing of Transition T_1 will force Transition T_2 to release its input bag. This causes a

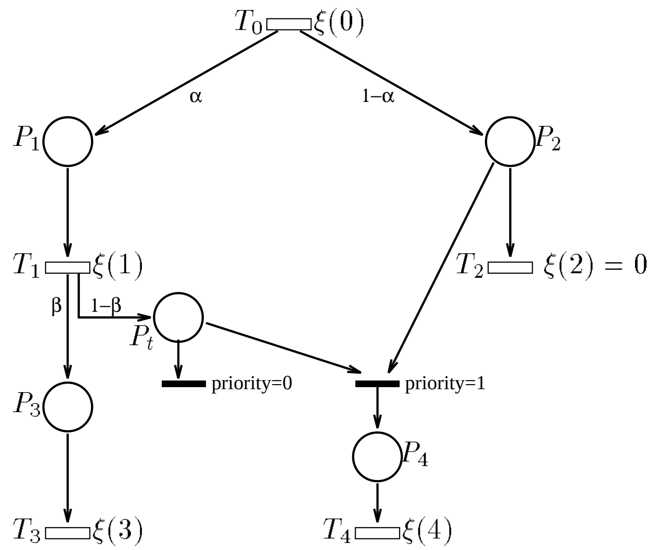

Figure 5.5: Simulated net using conflict resolution

single token to move from P_2 to P_4, if one is present. If no token is present in P_2 when the forcing is attempted then the net effect on the SPN is that the token, which was released from P_1 and deposited in P_t, is lost. The tokens that build up at P_3 and P_4 are removed from the SPN when transitions T_3 and T_4 fire respectively (each of these transitions has $\mathbf{0}$ as their output bag).

The Exact Solution

Although it is not readily apparent that the SPN in Figure 5.5 has a product form equilibrium distribution we can apply Theorem 5.2.1. We do not consider place P_t in this analysis as it is a temporary place, included only to facilitate in the description of forced transition firing. In the following we label all variables by the corresponding transition's index in the intuitive way. The state of the marking process is $\mathbf{m} = (m_1, m_2, m_3, m_4)$, and $\mathcal{A} = \{\mathbf{0}, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$. The input bag of transition T_i is $\mathbf{e}_i$, for $i = 1, 2, 3, 4$, and $I(0) = \mathbf{0}$. For this SPN equations (5.7) are

$$\mathbf{0} \cdot \mathbf{x} = 0 = log(f(0)), \; i = 1, 2, 3, 4, \quad \text{and} \quad \mathbf{e}_i \cdot \mathbf{x} = x_i = log(f(i)).$$

Thus $f(0) = 1$ and equations (5.6) become

$$\xi(1) f(1) \quad = \quad \xi(0)\alpha,$$

$$
\begin{aligned}
0 &= \xi(0)(1-\alpha) - \xi(1)f(1)f(2)(1-\beta), \\
\xi(3)f(3) &= \xi(1)f(1)\beta, \\
\xi(4)f(4) &= \xi(1)f(1)f(2)(1-\beta),
\end{aligned}
$$

for which the unique solutions are

$$
\begin{aligned}
f(1) &= \frac{\xi(0)\alpha}{\xi(1)}, \\
f(2) &= \frac{1-\alpha}{\alpha(1-\beta)}, \\
f(3) &= \frac{\xi(0)\alpha\beta}{\xi(3)}, \\
f(4) &= \frac{\xi(0)(1-\alpha)}{\xi(4)}.
\end{aligned}
\tag{5.18}
$$

Thus, assuming $\xi(0) > 0$, $0 < \alpha < 1$, and $0 < \beta < 1$, there exists a solution $\{f(\mathbf{a}) > 0,\ \forall \mathbf{a} \in \mathcal{A},\ \mathbf{x}\}$ to equations (5.6) and (5.7). In this example, $f(i)$ is the probability that transition T_i is enabled. The equilibrium distribution can then be calculated from equations (5.18) using equations (5.7) and (5.8). It is

$$\pi(\mathbf{n}) = K \prod_{i=1}^{4} f(i)^{m_i}.$$

A Numerical Comparison

As an additional check we wish to compare our results against those obtained using the package SPNP [14] which simulates closed Petri nets. However the SPN in Figure 5.5 is an open SPN and must be approximated, when using SPNP, by the addition of an environment place to represent the outside of the SPN, containing K tokens in the initial marking For our numerical analysis we set $\xi(0) = 1$, $\xi(1) = 2$, $\xi(2) = 0$, $\xi(3) = \frac{1}{5}$, $\xi(4) = 1$, $\alpha = \frac{3}{4}$, and $\beta = \frac{1}{5}$, so that from equations (5.18) we expect $f(0) = 1$, $f(1) = \frac{3}{8}$, $f(2) = \frac{5}{12}$, $f(3) = \frac{3}{4}$, and $f(4) = \frac{1}{4}$. Table 5.1 shows how the state space increases rapidly with increasing K and therefore how much easier it is to analyse these classes of SPNs when they are known to have product form equilibrium distributions. The required CPU time increased linearly with the size of the state space and exponentially with the number of transitions. We do not have results from the simulation for $f(2)$, as SPNP does not model transitions with zero firing rates. Observe that the SPNP results are converging to the results established by Theorem 5.2.1.

Transition	K=5	K=10	K=15	K=20	K=25	Exact
T0	0.78126	0.95527	0.99003	0.99768	0.99945	1
T1	0.29405	0.35824	0.37126	0.37413	0.37479	0.375
T3	0.58811	0.71649	0.74252	0.74826	0.74959	0.75
T4	0.19315	0.23879	0.24751	0.24942	0.24986	0.25
States	79	580	2159	5785	12739	

Table 5.1: Simulated and Exact Transition Enabling Probabilities

We note that there are other methods, from the Petri net literature, for drawing SPNs with forced transition firing. One such method is the use of inhibitor arcs. However, the number of transitions needed to describe conflict resolution exhaustively using inhibitor arcs becomes very large when the conflicting transitions have multiple places in their respective input bags. This prohibits the simulation of any but the simplest SPNs, with forced transition firing, drawn using inhibitor arcs, as the required CPU time increases exponentially with the number of transitions.

5.3.2 Resource Allocation Problems

Consider a system consisting of a Jackson network of queues connected to a SPN. The tokens in the SPN represent resources available for use by the customers which move around in the Jackson network. Resources may be replenished by arriving from outside the system to an open SPN or, alternatively, there may be a fixed number of resources, in which case the SPN is closed. Tokens may move around the SPN only when they are accessed by customers from the Jackson network. Customers may be thought of as jobs, or tasks, that require processing before they access the resources. This may involve, for example, prioritising or scheduling of the tasks, and the tasks must compete for the available resources. A task accesses the resources when its corresponding customer leaves the queueing network and attempts to force the release of a vector of resources in the SPN. The forced release of resources will be successful only if the resources are available. Once a task is complete, or stalled, the customer corresponding to that task will leave the system entirely. A stalled task is one in which the resources are not available for completion of the task. A single task may require multiple accesses to the resources, and so, after each successful access, a customer may return to the queueing network for more processing.

Resources arrive to, and are stored at, the SPN. We do not allow transitions to fire within the SPN. The conditions in Theorem 5.2.3 need to be satisfied for the system, which is the union of the queueing network and the SPN. Since transitions may not fire within the SPN, the conditions require that any vector deposited in the SPN be forcibly released, at some stage, by a customer from the queueing network. Customers arriving to the SPN will always attempt to force the release of a vector of resources, and will do one (and only one) of the following:

- return to the queueing network as a single customer (the resources have been consumed),
- deposit a vector of resources in the SPN (the resources have been processed and altered),
- leave the system entirely (the task is complete or stalled).

We model this by considering a system of N nodes. Nodes labelled $1, 2, \ldots, N_R$ are "resource" places in the SPN and $N_R + 1, \ldots, N$ are "customer" queues in the queueing network.

Let entities have a subscript "r" and "c" when referring to the resource SPN and customer queueing network respectively, and no subscript when referring to the entire system. The behavioural constraints on the system are modelled by letting $\mathbf{n}$ =

$(\mathbf{n}_r, \mathbf{n}_c)$, and $\mathbf{a} = (\mathbf{a}_r, \mathbf{a}_c)$, restricting $\mathbf{a}_c$ to be either $\mathbf{0}$ or $\mathbf{e}_i$, where $\mathbf{e}_i$ is a vector of zeros with a 1 as the i-th entry. Set

$$\xi(\mathbf{a}) = \begin{cases} \mu_0, & \text{if } \mathbf{a} = \mathbf{0}, \\ \mu_i, & \text{if } \mathbf{a} = (\mathbf{0}, \mathbf{e}_i), \\ 0, & \text{otherwise.} \end{cases}$$

The transition probabilities of the system are restricted by setting

$$p(\mathbf{a}, \mathbf{a}', \mathbf{a}'') = 0 \text{ if } \begin{cases} \mathbf{a}_r \neq \mathbf{0}, \\ \mathbf{a}'_c \neq \mathbf{0}, \\ (\mathbf{a}''_r \neq \mathbf{0}) \text{ and } (\mathbf{a}''_c \neq \mathbf{0}), \\ (\mathbf{a} \neq \mathbf{0}) \text{ and } (\mathbf{a}' = \mathbf{0}) \text{ and } (\mathbf{a}''_r \neq \mathbf{0}), \\ (\mathbf{a} = \mathbf{0}) \text{ and } (\mathbf{a}' \neq \mathbf{0}), \end{cases}$$

so that the customer and resource partitions communicate as described. Also set

$$\psi(\mathbf{n}) = \phi(\mathbf{n}) = \begin{cases} \prod_{i=N_R+1}^{N} \prod_{k=0}^{n_{ci}} \nu_i(k)^{-1}, & \text{if } \mathbf{n} \geq \mathbf{0}, \\ 0, & \text{otherwise,} \end{cases}$$

where n_{ci} is the i-th component of $\mathbf{n}_c$. The state space of the system is implicitly defined by this equation to be

$$S = \{\mathbf{n} : \mathbf{n} \geq \mathbf{0}\},$$

since $\pi(\mathbf{n}) = 0$, in equation (5.8), if any component of $\mathbf{n}$ is negative. Thus equation (5.1) becomes

$$q(\mathbf{n}, \mathbf{a}) = \begin{cases} \mu_0, & \text{if } \mathbf{a} = (\mathbf{0}, \mathbf{0}), \\ \mu_i \nu_i(n_{ci}), & \text{if } \mathbf{a} = (\mathbf{0}, \mathbf{e}_i), \\ 0, & \text{otherwise,} \end{cases}$$

and when the system is in state $\mathbf{n} \in S$ the forced release of resources $\mathbf{a}_r$ by a customer emission from queue i in the queueing network will be successful with probability, given by equation (5.3) to be

$$\frac{\psi((\mathbf{n}_r - \mathbf{a}_r, \mathbf{n}_c - \mathbf{e}_i))}{\psi((\mathbf{n}_r, \mathbf{n}_c - \mathbf{e}_i))}.$$

With the above definition of $\psi(\mathbf{n})$ this probability will be one if $\mathbf{n}_r \geq \mathbf{a}_r$, otherwise the probability will be zero, implying that the corresponding task will be stalled due to insufficient resources being available.

Customers are served at customer queues at state dependent rates, given by the functions $\nu_i(\cdot)$ and can move between customer queues as they would in a standard

Jackson network. However, occasionally a customer from queue i will attempt to access a vector of resources, $\mathbf{a}_r$, in the SPN and deposit a vector, $\mathbf{a}''$, in the system with intensity $\mu_i \nu_i(n_{ci}) p((\mathbf{0}, \mathbf{e}_i), (\mathbf{a}_r, \mathbf{0}), \mathbf{a}'')$. If the resources are not available in the SPN the customer released from queue i leaves the network. When a customer successfully accesses a vector of resources it may process them and immediately deposit a different vector of resources, it may consume the resources and return to the customer queueing network to complete other tasks, or it may leave the network entirely with the resources that it has appropriated. This system satisfies the conditions of Theorem 5.2.3, but we have not determined whether there exists a positive solution $\{f(\mathbf{a}),\ \mathbf{a} \in \mathcal{A}\}$ to equations (5.6). We will do so, in the next section, for loss networks, which we model using the resource allocation framework presented here.

5.3.3 Loss Networks

We use the resource allocation model to describe a system in which origin-destination traffic is carried by a loss network (LN). This enables us to determine deficiencies for the system under different traffic loads, and model circuit availability within the LN. A typical LN consists of a set of N nodes connected by a number, L say, of links. Link i contains c_i circuits, for $i = 1, 2, \ldots, L$. An example of a LN is shown in Figure 5.6.

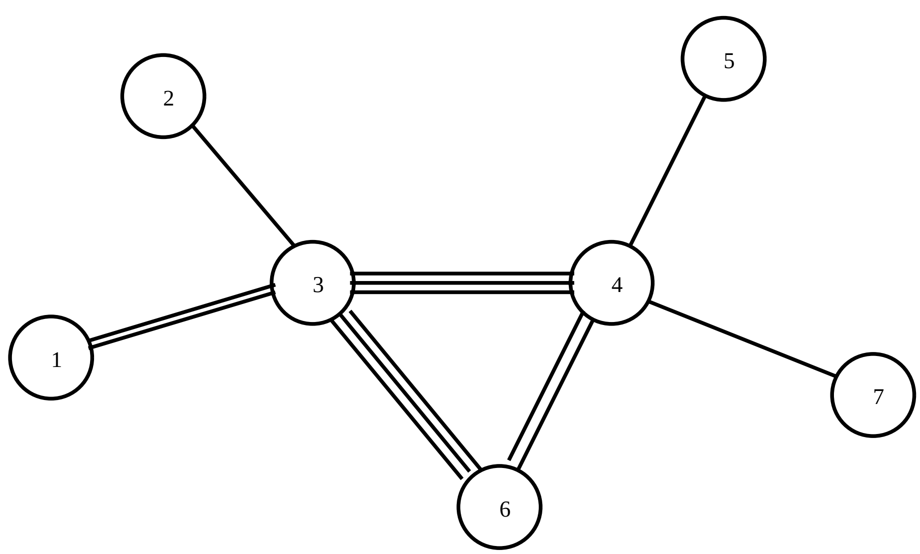

Figure 5.6: A typical loss network

The number of circuits available for use (the configuration of available resources in the LN) is described by the vector $\mathbf{m} = (m_1, m_2, \ldots, m_L)$, where m_i is the number of circuits currently available on link i. When no circuits are being used $\mathbf{m} = \mathbf{c} = (c_1, c_2, \ldots, c_L)$. The LN is capable of carrying different types of traffic between each origin-destination pair of nodes. For example, it is common place to model voice traffic as requiring only one circuit on each link along the route connecting its origin and destination, and video traffic as requiring more than one circuit per link. There are a finite number of origin-destination pairs, a finite number of routes between them, and a finite number of traffic types to be carried. Let each possible combination of origin-destination pairs, routes and traffic types be represented by a unique call type. Let the total number of call types be T. $\mathcal{A} = \{\mathbf{a}_t,\ t = 1, 2, \ldots, T\}$ is the set of elements that describes the number of circuits required, on each link in the LN, for each call type. $\mathbf{a}_t = (a_{t1}, a_{t2}, \ldots, a_{tL})$, where a_{ti} is the number of circuits required on link i by a type t call. To establish connectivity for a type t call, and carry traffic, the LN must have a configuration $\mathbf{m} \geq \mathbf{a}_t$ of available circuits. If the call is established then $\mathbf{a}_t$ circuits must be reserved, so that the new configuration of available circuits is $\mathbf{m} - \mathbf{a}_t$. If $m_i < a_{ti}$ for any link in the LN, then type t calls will be blocked. Type t calls will continue to be blocked until $\mathbf{m} \geq \mathbf{a}_t$. The circuits become available again once the call is finished.

The resource allocation model is represented by Figure 5.7. We use a closed SPN consisting of L isolated nodes, one for each link, labelled $1r, 2r, \ldots, Lr$, to model the configuration of resources currently available in the LN. The processes corresponding to the arrival, connection, and holding times for calls are modelled using a queueing network, which is attached to the SPN as in the resource allocation model in Section 5.3.2. The queueing network itself is divided into two disjoint sections, a call processing section (section P), and a call holding section (section H). The nodes of the queueing network are labelled by the set $\mathcal{L} = \{1p, 2p, \ldots, Tp, 1h, 2h, \ldots, Th\}$.

Let calls of each type arrive to the network in independent Poisson streams, with

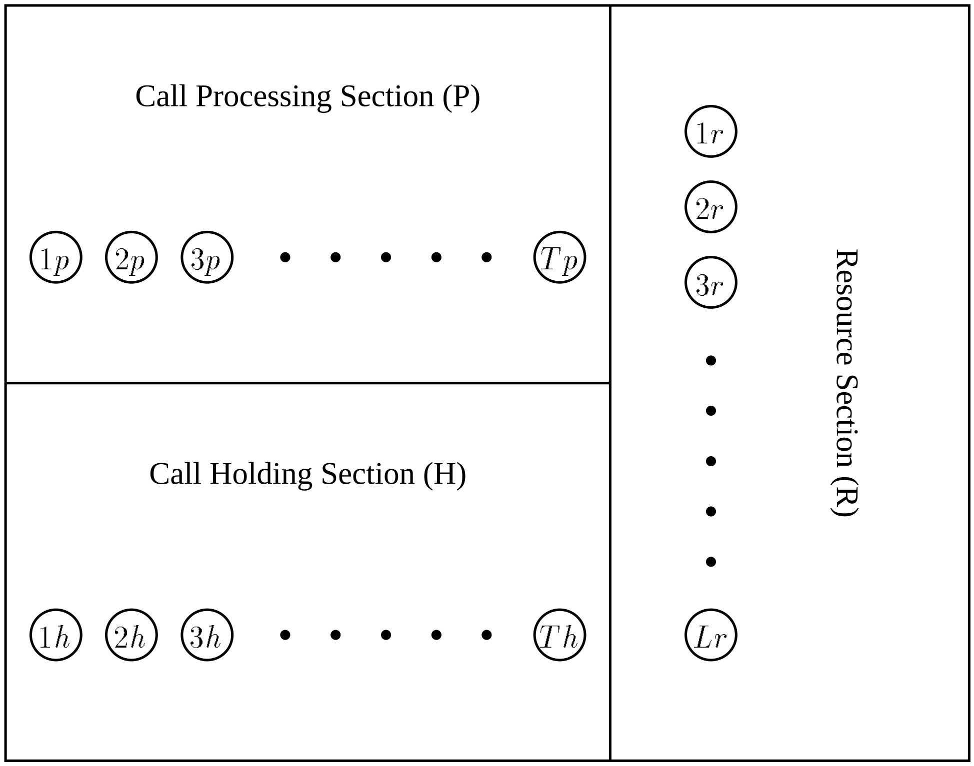

Figure 5.7: Resource allocation model of a loss network

calls of type t queueing at node tp to be processed. Calls do not arrive from outside the network to nodes $th,\ t = 1, 2, \ldots, T$. Node th is used as the holding place for calls of type t in progress. Having been processed at node t a customer, who corresponds to a type t call, will attempt to remove a vector, $\mathbf{a}_t$, of circuits from the SPN, and subsequently deposit a customer at node th. If there are not enough circuits available then the customer is lost from the system, so that the call is blocked. If the circuits are available, then they are reserved for the duration of the call. The number of customers queued at node th, n_{th}, represents the number of type t calls in progress. Customers arriving to node th immediately start receiving service, with the service time corresponding to the call holding time. We model this by making each node in section H an infinite server queue (the nodes in section P have single servers operating under the FCFS service discipline). Once the call is complete, the customer is released from node th, transformed into a vector of circuits, $\mathbf{a}_t$, and deposited in the SPN. The circuits are then available for use by other calls.

The number of circuits on each link is fixed and so

$$\mathbf{c} = \mathbf{m} + \sum_{t=1}^{T} n_{th}\mathbf{a}_t, \tag{5.19}$$

where $\mathbf{m}$ is the marking of the SPN. It is important to note that, once $\mathbf{n}_h = (n_{1h}, n_{2h}, \ldots, n_{Th})$ is known, we can determine $\mathbf{m}$ from equation (5.19). Therefore we have full information about the system when we know the state of the queueing network, namely

$$\mathbf{n} = (n_{1p}, \ldots, n_{Tp}, n_{1h}, \ldots, n_{Th}) = (\mathbf{n}_p, \mathbf{n}_h).$$

Thus the equilibrium distribution, for the whole system, depends only on $\mathbf{n}$, and not the augmented state vector $(\mathbf{n}, \mathbf{m})$.

Using a compact notation, where a unit vector corresponds to a node in the queueing network, and a vector $\mathbf{a}_i \in \mathcal{A}$ corresponds to the set of links required by a type i call, we set

$$\xi(\mathbf{a}) = \begin{cases} \mu_0, & \text{if } \mathbf{a} = \mathbf{0}, \\ \mu_i, & \text{if } \mathbf{a} = \mathbf{e}_i,\ i \in \mathcal{L} \\ 0. & \text{otherwise,} \end{cases}$$

The service disciplines of the nodes in the queueing network are modelled by setting

$$\nu_i(n) = \begin{cases} n, & \text{if } n > 0,\ i = 1h, 2h, \ldots, Th, \\ 1, & \text{otherwise,} \end{cases}$$

and set the transition probabilities, for $i \in \{1, 2, \ldots, T\}$, $\mathbf{a}_i \in \mathcal{A}$, to be

$$\begin{aligned} p(\mathbf{0}, \mathbf{0}, \mathbf{e}_i) &= 1, \\ p(\mathbf{e}_i, \mathbf{a}_i, \mathbf{e}_{i0}) &= 1, \\ p(\mathbf{e}_{i0}, \mathbf{0}, \mathbf{a}_i) &= 1, \end{aligned}$$

and zero for any other $p(\mathbf{a}, \mathbf{a}', \mathbf{a}'')$. The traffic equations (5.6) simplify considerably to become, for $i \in \{1, 2, \ldots, T\}$,

$$\mu_{ip} f(\mathbf{e}_{ip}) = \mu_0, \tag{5.20}$$

$$\begin{aligned} \mu_{ih} f(\mathbf{e}_{ih}) &= \mu_{ip} f(\mathbf{e}_{ip}) f(\mathbf{a}_i), \\ &= \mu_0 f(\mathbf{a}_i), \end{aligned} \tag{5.21}$$

and, with

$$\mathbf{x} = (x_{1p}, \ldots, x_{Tp}, x_{1h}, \ldots, x_{Th}) = (\mathbf{x}_p, \mathbf{x}_h),$$

equations (5.7) become

$$x_{ip} = log(f(\mathbf{e}_{ip})), \quad i \in \{1, 2, \ldots, T\}, \tag{5.22}$$

$$x_{ih} = log(f(\mathbf{e}_{ih})), \quad i \in \{1, 2, \ldots, T\}. \tag{5.23}$$

Obviously there is redundancy equation (5.21), involving $f(\mathbf{e}_{ih})$ and $f(\mathbf{a}_i)$. This occurs because of the following. We may consider the amalgamation of section H and section R to be a closed process. Exogenous events, namely the release of a customer from section P, determine when resources are released from section R, resulting in a customer being placed in section H. The traffic equations for a closed system always exhibit redundancies. The problem is alleviated by noting that $f(\mathbf{a}_i)$, for $i = 1, 2, \ldots, T$, does not appear in equations (5.22) and (5.23), and so will not appear in the invariant measure. Therefore we may set $f(\mathbf{a}_i)$ to be a constant, for $i = 1, 2, \ldots, T$. It is convenient to set them all to be one. Therefore

$$f(\mathbf{e}_{ip}) = \frac{\mu_0}{\mu_{ip}}, \tag{5.24}$$

$$f(\mathbf{e}_{ih}) = \frac{\mu_0}{\mu_{ih}}, \tag{5.25}$$

and thus $\mathbf{x}$ can be determined uniquely. The resulting invariant measure is

$$exp(\mathbf{n} \cdot \mathbf{x}) = \prod_{i=1}^{T} \left(\frac{\mu_0}{\mu_{ip}} \right)^{n_{ip}} \prod_{j=1}^{T} \left(\frac{\mu_0}{\mu_{jh}} \right)^{n_{jh}}.$$

Further complexity may be added to the model if we wish. For example, to model delays due to communications protocols, each call may be made to route through a series of additional nodes in section P of the queueing network. Calls of different types are allowed to have different service rates at each node, so that they do not have to satisfy the same set of protocols. This entire model is a special case of the resource allocation models in Section 5.3.2, which are themselves a special case of the queueing networks with forced batch movement presented in this chapter.

Concluding Remarks

Summary

In this thesis we have extended the literature on signalling in product form queueing networks. In doing so we have also combined a number of previous results.

The queueing networks of Chapter 3 allow signalling entities to behave as triggers or batches of negative customers. The arrival and service intensities of these networks are state dependent which, among other things, allows for state space truncation, multiple customer classes, and the ability for a node to ignore an arriving signal. The state dependence, arising from the use of the functions $\psi(\cdot)$ and $\phi(\cdot)$, effectively imposes speeds on the network, which may vary between states.

The large amount of attention that queueing networks with negative customers, triggers and signals have received in the literature recently is justified by the ease with which they lend themselves to a number of modelling applications. We have demonstrated a few of theses applications, in areas as diverse as manufacturing processes, predator-prey models, and communications protocols.

In Chapter 4 we have identified the way in which networks, incorporating signalling transitions, may be considered to be derived from a coupling/linking of quasi-reversible nodes. By definition, a node with state dependent intensities may not be quasi-reversible, implying that we are not able to truncate the state space of a quasi-reversible node using $\psi(\cdot)$ and $\phi(\cdot)$ functions. For this reason we have focussed attention on processes with non-negative queue lengths in the discussion on quasi-reversibility. The coupling together of quasi-reversible nodes to obtain a quasi-reversible network which is equivalent to one with triggering transitions is complicated by the fact that the final deposit of a customer in a triggered transition relies on the intermediary node being non-empty.

Recently a number of researchers have identified a relationship between single movement queueing networks and batch movement queueing networks. We have used this relationship in Chapter 5 to derive a network of queues with forced batch movement. Special conditions must hold for the resulting batch movement network to have a product form, one of which is a release-deposit condition. We have applied the batch movement queueing network to the modelling framework of SPNs to obtain a new class of SPN in which a transition may be forced to fire before the lifetime for that transition has expired. Finally, we have proposed a modelling framework for resource allocation problems, and have used it to model a loss network.

Further Work

The possible generalisation of coupling techniques to include more than two nodes in a single transition is an area of research worthy of attention. If it is possible to determine a general method of coupling nodes in this way, then a whole new class of product form queueing networks will be established. Batch movement queueing networks may also be generalised further by introducing all types of signalling (not just triggers) in the one batch movement queueing network. This may place too large a restriction on the routeing chain however.

Bibliography

[1] Balsamo, S., and Iazeolla, G., *An extension of Norton's theorem for queueing networks*, IEEE Soft. End., **8**, pp 298–305, (1982).

[2] Balsamo, S., and Iazeolla, G., *Aggregation and disaggregation in queueing networks : the principle of product form synthesis*, Mathematical Computer Performance Reliability, Iazeolla, Courtois and Hordijk (editors), Elsener Science Publishers B.V., North Holland, pp 95–109, (1984).

[3] Baskett F., Chandy K., Muntz R., and Palacios J., *Open, closed and mixed networks of queues with different classes of customers*, J. Assoc. Comput. Mach., **22**, pp 248–260, (1975).

[4] Boucherie, R. and van Dijk, N., *Spatial birth-death processes with multiple changes and applications to batch service networks and clustering processes*, Adv. Appl. Prob., **22**, pp 433–455, (1990).

[5] Boucherie, R. and van Dijk, N., *Product forms for queueing networks with state dependent multiple job transitions*, Adv. Appl. Prob., **23**, pp 152–187, (1990).

[6] Boucherie, R. and van Dijk, N., *Local balance in queueing networks with negative customers*, to appear in Ann. Ops. Res., (1993).

[7] Burke, P.J., *The output of a queueing system*, Operations Research, **4**, p 699, (1956).

[8] Buzen, J., *Queueing network models of multiprogramming*, PhD thesis, Harvard U., Cambridge, Mass., (1971).

[9] Buzen, J., *Computational algorithms for closed queueing networks with exponential services*, Comm., A.C.M., **16**, pp 527–531, (1973).

[10] Chandy, K.M., Herzog, W., and Woo, L., *Parametric analysis of queueing networks*, IBM J. Res. Develop., **19**, pp 36–42, (1975).

[11] Chandy, K.M., Herzog, W., and Woo, L., *Approximate analysis of general queueing networks*, IBM J. Res. Develop., **19**, pp 43–49, (1975).

[12] Chandy, K.M., Howard, J.H., and Towsley, D.F., *Product form and local balance in queueing networks*, J. Assoc. Comput. Mach., **24**, pp 250–263, (1977).

[13] Chandy, K.M., and Martin, A.J., *A characterisation of product form queueing networks*, J. Assoc. Comput. Mach., **30**, pp 286–299, (1983).

[14] Ciardo, G., Muppala, J., and Trivedi, K.S., *SPNP: Stochastic Petri net package*, Proc. 3^{rd} International Workshop on Petri Nets and Performance Models, (IEEE Computer Society Press), pp 142–151, (1989).

[15] Ciardo, G., and Trivedi, K.S., *A decomposition approach for stochastic Petri net models*, Proc. Petri nets and performance models, Melbourne, pp 74–85, (1991).

[16] Coleman, J.L., Henderson, W., and Taylor, P.G., *A Convolution Algorithm for Calculating Exact Equilibrium Distributions in Resource Allocation Problems with Moderate User Interference,* to appear in IEEE Trans. Comm., (1993).

[17] Coleman, J.L., Henderson, W., and Taylor, P.G., *Product form equilibrium distributions and a convolution algorithm for stochastic Petri nets,* Submitted, (1993).

[18] Coleman, J., Henderson, W., and Taylor, P.G., *Algorithms for evaluating the normalising constant in stochastic Petri nets with product form equilibrium distributions*, submitted, (1993).

[19] Coleman, J., Henderson, W., Pearce, C.E.M., and Taylor, P.G., *Mean value analysis for batch movement queueing networks with product form equilibrium distributions*, submitted, (1993).

[20] Coyle, A.J., Henderson, W., and Taylor, P.G., *Reduced load approximations for loss networks*, to appear in Telecommunications Systems, (1993).

[21] Donatelli, S., and Sereno, M., *On the product form solution for stochastic Petri nets,* Proc. 13^{th} International Conference on Application and Theory of Petri Nets, Sheffield UK, pp 154–172, (1992).

[22] Dunford, N. and Schwartz, N.J. *Linear Operators part 1*, p 453, proof p 467, Interscience Publishers, New York, (1958-71).

[23] Forneau, J.M., and Gelenbe, E. *Multiple class G-networks*, O.R.S.A. Technical committee on Computer Science Conference, Williamsburg, (1992).

[24] Frosch, D., and Natarajan, K., *Product Form Solutions for Closed Synchronised Systems of Stochastic Sequential Processes,* International Computer Symposium, Taichung, Taiwan, (1992).

[25] Gelenbe, E., *Random neural networks with negative and positive signals and product form solution* Neural Comp., **1**, pp 502–510, (1989).

[26] Gelenbe, E., *Stability of the random neural network model*, Neural Comp., **2**, pp 239–247, (1990).

[27] Gelenbe, E., *Product form networks with negative and positive customers*, J. Appl. Prob., **28**, pp 656–663, (1991).

[28] Gelenbe, E., Glynn, P., Sigman, K., *Queues with negative arrivals*, J. Appl. Prob., **28**, pp 245–250, (1991).

[29] Gelenbe, E., *G-networks with triggered customer movement*, preprint, (1992).

[30] Gelenbe, E., *Negative customers with batch removal*, personal communication, (1992).

[31] Gelenbe, E., and Schassberger, R., *Stability of G-networks*, Probability and its applications in engineering and the information sciences, Cambridge University Press, (1992).

[32] Gordon, W.J., and Newell, G.F., *Closed queueing systems with exponential servers*, Operat. Res., **15**, pp 254–265, (1967).

[33] Henderson, W. and Taylor, P.G., *Insensitivity of processes with interruptions* J. Appl. Prob., **26**, pp 242–258, (1989).

[34] Henderson, W., Lucic, D., and Taylor, P.G., *A net level performance analysis of stochastic Petri nets* J. Aust. Math. Soc. Ser. B, **31**, pp 176–187, (1989).

[35] Henderson, W., Pearce, C.E.M., Taylor, P.G., and Van Dijk, N.M., *Closed Queueing networks with batch services,* Queueing Systems, **6**, pp 59–70, (1990).

[36] Henderson, W., and Taylor, P.G., *Product form in networks of queues with batch arrival and batch services*, Queueing Systems, **6**, pp 71–88, (1990).

[37] Henderson, W., and Taylor, P.G., *Embedded processes in stochastic Petri nets*, IEEE Transaction on software engineering, **17**, pp 108–116, (1991).

[38] Henderson, W., and Taylor, P.G., *Some new results on queueing networks with batch movement,* J. Appl. Prob., **28**, pp 409–421, (1991).

[39] Henderson W., Pearce C.E.M., Pollett P.K., and Taylor P.G., *Connecting internally-balanced quasi-reversible processes*, Adv. in Appl. Prob., **24**, pp 934–959, (1992)

[40] Henderson, W., Northcote, B.S., and Taylor, P.G. *Geometric equilibrium distributions for queues with interactive batch departures*, to appear in Ann. Ops. Res., **48**(1), (1994).

[41] Henderson, W., *Queueing networks with negative customers and negative queue lengths*, to appear in J. Appl. Prob., **30**(3), (1993).

[42] Henderson, W., Northcote, B.S., and Taylor, P.G. *State dependent signalling in queueing networks*, to appear in Adv. Appl. Prob., **26**(2), (1994).

[43] Henderson, W., Northcote, B.S., and Taylor, P.G. *Product form in networks of queues with forced batch movement*, submitted, (1993).

[44] Henderson W., Pearce C.E.M., Schassberger R. and Taylor P.G. *Product form in Markov processes*, submitted, (1993).

[45] Hunt, G.C., *Sequential arrays of waiting lines*, Operations Research, **4**, p 674, (1956).

[46] Jackson, J.R., *Networks of waiting lines*, Operat. Res., **5**, pp 518–521, (1957).

[47] Jackson, J.R., *Jobshop-like queueing systems*, Management Science, **10**, pp 131–142, (1963).

[48] Kandel, E., and Schwartz, J., *Principles of Neural Science*, Elsevier, Amsterdam, (1985).

[49] Kelly F.P., *Networks of queues with customers of different types*, J. Appl. Prob., **12**, pp 542–554, (1975).

[50] Kelly F.P., *Markov processes and Markov random fields* Bull. Inst. Int. Statist., **46**, pp 397–404, (1975).

[51] Kelly F.P., *Networks of Queues*, Adv. Appl. Prob., **8**, pp 416–432, (1976).

[52] Kelly, F.P., *Reversibility and stochastic networks*, Wiley, New York, (1979).

[53] Kelly, F.P., *Networks of quasi-reversible nodes*, Proc. ORSA-TIMS Special Interest Meeting : Applied Probability – Computer Science, the Interface, Birkhäuser-Boston, Cambridge, Mass., (1981).

[54] Kingman J.F.C., *Markov population processes*, J. Appl. Prob., **6**, pp 1–18, (1969).

[55] Miller, *Stationary equations in continuous time Markov chains*, Trans. American Math. Soc., **109**, pp 35–44, (1963).

[56] Li, M., and Georganas, N.D., *Parametric analysis of stochastic Petri nets*, 5^{th} international conference on modelling and tools for computer performance evaluation, Torino, Italy, (1991).

[57] Miyazawa, M., *Insensitivity and product form decomposability of reallocatable GSMP*, submitted to Adv. Appl. Prob., (1992).

[58] Muntz, R.R., *Poisson departure processes and queueing networks*, IBM Research Report RC4145, IBM Thomas J. Watson Research Centre, Yorktown Heights, New York. A shortened version appears in Proc. 7^{th} annual Conference on Information Science and Systems, Princeton, pp 435–440, (1973).

[59] Peterson, J.L., *Petri Net Theory and the Modeling of Systems,* Englewood Cliffs, NJ: Prentice-Hall, (1981).

[60] Pollett P.K., *Connecting reversible processes*, Adv. Appl. Prob., **18**, pp 880–900, (1986).

[61] Pollett P.K., *Preserving partial balance in continuous-time Markov chains*, Adv. Appl. Prob., **18**, pp 431–453, (1987).

[62] Reich, E., *Birth-death processes and tandem queues*, Rand Corporation, p 863, (1956).

[63] Reiser, M., and Lavenberg, S.S., *Mean-value analysis of closed multi-chain queueing networks,* J. Assoc. Comput. Mach., **27**, pp 313–322, (1980).

[64] Ross, Sheldon M., *Introduction to probability models*, 4th edition, Academic Press, (1989).

[65] Schassberger R., *Insensitivity of steady-state distributions of generalised semi-Markov processes: Part I*, Ann. Prob., **5**, pp 87–99, (1977).

[66] Schassberger R., *Insensitivity of steady-state distributions of generalised semi-Markov processes: Part II*, Ann. Prob., **6**, pp 85–93, (1978).

[67] Schassberger R., *Insensitivity of steady-state distributions of generalised semi-Markov processes with speeds*, Adv. Appl. Prob., **10**, pp 836–851, (1978).

[68] Schassberger R., *The insensitivity of stationary probabilities in networks of queues*, Adv. Appl. Prob., **10**, pp 906–912, (1978).

[69] Schassberger R., *A definition of discrete product form distributions*, Z. Operat. Res., **23**, pp 189–195, (1979).

[70] Serfozo, R.F., *Markovian network processes with system dependent transition rates*, Georgia Institute of Technology.

[71] Walrand, J., and Varaiya, P., *Interconnections of Markov chains and quasi-reversible queueing networks*, Stoch. Process. Appl., **10**, pp 209–219, (1980).

[72] Walrand J., *A discrete time queueing network*, J. Appl. Prob., **20**, pp 903–909, (1983).

[73] Walrand J., *An Introduction to Queueing Networks*, Prentice Hall, New Jersey, (1988).

[74] Whittle, P., *Equilibrium distributions for an open migration process*, J. Appl. Prob., **5**, pp 567–571, (1968).

[75] Whittle, P., *Nonlinear migration process*, Bull. Internat. Inst. Statist., **42**, pp 642–647 (1969).

[76] Whittle, P., *Systems in Stochastic Equilibrium*, Wiley New York, (1986).